Springer Proceedings in Complexity

Alfredo J. Morales · Carlos Gershenson
Dan Braha · Ali A. Minai · Yaneer Bar-Yam
Editors

Unifying Themes in Complex Systems IX

Proceedings of the Ninth International Conference on Complex Systems

Springer

Springer Proceedings in Complexity

Springer Complexity

Springer Complexity is an interdisciplinary program publishing the best research and academic-level teaching on both fundamental and applied aspects of complex systems—cutting across all traditional disciplines of the natural and life sciences, engineering, economics, medicine, neuroscience, social, and computer science.

Complex Systems are systems that comprise many interacting parts with the ability to generate a new quality of macroscopic collective behavior the manifestations of which are the spontaneous formation of distinctive temporal, spatial, or functional structures. Models of such systems can be successfully mapped onto quite diverse "real-life" situations like the climate, the coherent emission of light from lasers, chemical reaction–diffusion systems, biological cellular networks, the dynamics of stock markets and of the Internet, earthquake statistics and prediction, freeway traffic, the human brain, or the formation of opinions in social systems, to name just some of the popular applications.

Although their scope and methodologies overlap somewhat, one can distinguish the following main concepts and tools: self-organization, nonlinear dynamics, synergetics, turbulence, dynamical systems, catastrophes, instabilities, stochastic processes, chaos, graphs and networks, cellular automata, adaptive systems, genetic algorithms, and computational intelligence.

The three major book publication platforms of the Springer Complexity program are the monograph series "Understanding Complex Systems" focusing on the various applications of complexity, the "Springer Series in Synergetics", which is devoted to the quantitative theoretical and methodological foundations, and the "SpringerBriefs in Complexity" which are concise and topical working reports, case-studies, surveys, essays, and lecture notes of relevance to the field. In addition to the books in these two core series, the program also incorporates individual titles ranging from textbooks to major reference works.

More information about this series at http://www.springer.com/series/11637

Alfredo J. Morales · Carlos Gershenson
Dan Braha · Ali A. Minai
Yaneer Bar-Yam
Editors

Unifying Themes in Complex Systems IX

Proceedings of the Ninth International Conference on Complex Systems

Springer

Editors
Alfredo J. Morales
New England Complex Systems Institute
Cambridge, MA, USA

Carlos Gershenson
Instituto de Investigaciones en Matemáticas Aplicadas y en Sistemas
Universidad Nacional Autónoma de México
Mexico, Distrito Federal, Mexico

Dan Braha
New England Complex Systems Institute and University of Massachusetts
Cambridge, MA, USA

Ali A. Minai
Department of Electrical Engineering and Computer Science
University of Cincinnati
Cincinnati, OH, USA

Yaneer Bar-Yam
New England Complex Systems Institute
Cambridge, MA, USA

ISSN 2213-8684 ISSN 2213-8692 (electronic)
Springer Proceedings in Complexity
ISBN 978-3-319-96660-1 ISBN 978-3-319-96661-8 (eBook)
https://doi.org/10.1007/978-3-319-96661-8

Library of Congress Control Number: 2018950772

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

New England Complex Systems Institute

President
Yaneer Bar-Yam
New England Complex Systems Institute
277 Broadway
Cambridge, MA 02143, USA

New England Complex Systems Institute (NECSI) is an independent academic research and educational institution with students, postdoctoral fellows, and faculty. In addition to the in-house research team, NECSI has co-faculty, students, and affiliates from MIT, Harvard, Brandeis, and other universities nationally and internationally.

NECSI has been instrumental in the development of complex systems science and its applications. We study how interactions within a system lead to its behavioral patterns, and how the system interacts with its environment. Our new tools overcome the limitations of classical approximations for the scientific study of complex systems, such as social organizations, biological organisms, and ecological communities. NECSI's unified mathematically based approach transcends the boundaries of physical, biological, and social sciences, as well as engineering, management, and medicine.

NECSI research advances fundamental science and its applications to real-world problems, including social policy matters. NECSI researchers study networks, agent-based modeling, multiscale analysis and complexity, data analytics, artificial intelligence, chaos and predictability, evolution, ecology, biodiversity, altruism, systems biology, cellular response, health care, systems engineering, negotiation, military conflict, ethnic violence, and international development.

NECSI conducts classes, seminars, and conferences to assist students, faculty, and professionals in their understanding of complex systems. NECSI sponsors postdoctoral fellows, provides research resources online, and hosts the International Conference on Complex Systems. Through its education, NECSI strives to contribute to science and the betterment of society.

New England Complex Systems Institute

Series Editor
Dan Braha, Alfredo J. Morales, Carlos Gershenson,
and Ali A. Minai
New England Complex Systems Institute
277 Broadway
Cambridge, MA 20143, USA

The world around us is full of the wonderful interplay of relationships and emergent behaviors. The beautiful and mysterious way that atoms form biological and social systems inspires us to new efforts in science. As our society becomes more concerned with how people are connected to each other than how they work independently, science has become interested in the nature of relationships and relatedness. Through relationships, elements act together to become systems, and systems achieve function and purpose. The elements act together to become systems, and systems achieve function and purpose. The study of complex systems is remarkable in the closeness of basic ideas and practical implications. Advances in our understanding of complex systems give new opportunities for insight in science and improvement of society. This is manifest in relevance to engineering, medicine, management, and education. We devote this book series to the communication of recent advances and reviews of revolutionary ideas and their application to practical concerns.

Introduction

The mysteries of highly complex systems that have puzzled scientists for years have finally been unraveling thanks to new analytical and simulation methods. A better understanding of concepts like complexity, emergence, evolution, adaptation, and self-organization has shown that seemingly unrelated disciplines have more in common than we thought. These fundamental insights require interdisciplinary collaboration that usually does not occur between academic departments. This was the vision behind the first International Conference on Complex Systems in 1997: not just to present research, but to introduce new perspectives and foster collaborations that would yield research in the future.

As more and more scientists began to realize the importance of exploring the unifying principles that govern all complex systems, the 2018 ICCS attracted a diverse group of participants representing a wide variety of disciplines. Topics ranged from economics to ecology, from physics to psychology, and from business to biology. Through pedagogical, breakout, workshop, and poster sessions, conference attendees shared discoveries that were significant both to their particular field of interest and to the general study of complex systems. These volumes contain the proceedings from that conference.

Even with the ninth ICCS, the science of complex systems is still in development. In order for complex systems science to fulfill its potential to provide a unifying framework for various disciplines, it is essential to provide a standard set of conventions to facilitate communication. This is another valuable function of the conference: It allowed an opportunity to develop a common foundation and language for the study of complex systems.

These efforts have produced a variety of new analytic and simulation techniques that have proven invaluable in the study of physical, biological, and socioeconomical systems. New methods of statistical analysis led to better understanding of polymer formation and complex fluid dynamics; further development of these methods has deepened our understanding of patterns and networks. The application of simulation techniques such as agent-based models, cellular automata, and Monte Carlo calculations to complex systems has increased our ability to understand and even predict behavior of systems which once seemed completely unpredictable.

The disruption of big data and machine intelligence enabled the observation of patterns of human behavior at unprecedented scales and a deeper understanding of the structure and dynamics of social and economical systems.

The concepts and tools of complex systems are of interest not only to scientists, but also to corporate managers, doctors, political scientists, and policy makers. The same rules that govern neural networks apply to social or corporate networks, and professionals have started to realize how valuable these concepts are to their individual fields. The International Conferences on Complex Systems have provided the opportunity for professionals to learn the basics of complex systems and share their real-world experience in applying these concepts.

Organization

Conference Emeritus

Yaneer Bar-Yam	New England Complex Systems Institute

General Chair

Ali A. Minai	University of Cincinnati

Program Chairs

Alfredo J. Morales	New England Complex Systems Institute
Dan Braha	New England Complex Systems Institute and University of Massachusetts Dartmouth
Carlos Gershenson	National Autonomous University of Mexico

Special Sessions Chairs

Marc Santolini	Northeastern University
Emma Towlson	Northeastern University

Sponsors Chair

Kim Severino	New England Complex Systems Institute

Logistics Chairs

Sageet Braha	New England Complex Systems Institute
Elisabeth Perry-Montgomery	New England Complex Systems Institute

Program Committee

Fatimah Abdul Razak	Universiti Kebangsaan Malaysia
Andres Abeliuk	University of Southern California
Eran Agmon	Stanford University
Luca Maria Aiello	Nokia Bell Labs
Ma Al-Mamun	Yale University
Sergio Antonio Alcala-Corona	National Institute of Genomic Medicine, Mexico
Simone Alfarano	University Jaume I of Castellon
Eduardo Altmann	School of Mathematics and Statistics, The University of Sydney, Australia
Yaniv Altshuler	Massachusetts Institute of Technology
Alberto Antonioni	Carlos III University of Madrid
Alex Arenas	Universitat Rovira i Virgili
Eugenio Azpeitia	University of Zurich
Andres Baeza	Arizona State University
James Bagrow	University of Vermont
Selim Balcisoy	Sabanci University
Jose Balsa	Massachusetts Institute of Technology
Yaneer Bar-Yam	New England Complex Systems Institute
Federico Bassi	Université de Nantes
Manuel Bedia	University of Zaragoza
Rahmatollah Beheshti	Johns Hopkins University
Rosa M. Benito	Universidad Politécnica de Madrid
Katie Bentley	Boston University
Ginestra Bianconi	Queen Mary University of London
Philippe Binder	University of Hawaii at Hilo
Riccardo Boero	Los Alamos National Laboratory
Johan Bollen	Indiana University Bloomington
Denis Boyer	Universidad Nacional Autonoma de Mexico
Burcin Bozkaya	Sabanci University
Dan Braha	New England Complex Systems Institute
Jane Bromley	The Open University
Alessio Cardillo	Catalan Institute of Human Paleoecology and Social Evolution
Jeffrey Cares	Alidade Incorporated

Brian Castellani	Kent State University
Claudio Castellano	SMC, INFM-CNR
Myong-Hun Chang	Cleveland State University
Yu Chen	Binghamton University
Changqing Cheng	Binghamton University
Kenneth Chiu	Binghamton University
Philip Chodrow	Massachusetts Institute of Technology
Chun-An Chou	Northeastern University
Sean Cornelius	Northeastern University
Catherine Cramer	Columbia University
Herbert Dawid	Bielefeld University
Manlio De Domenico	Fondazione Bruno Kessler
Guillaume Deffuant	Cemagref
Domenico Delli Gatti	Università Cattolica
James Dixon	University of Connecticut
Xiaowen Dong	Massachusetts Institute of Technology
Fábio Duarte	Massachusetts Institute of Technology
Victor M. Eguiluz	Universitat de les Illes Balears
Kenth Engø-Monsen	Telenor ASA
Irving Epstein	Brandeis University
Tim Evans	Imperial College London
Paul Expert	Imperial College London
Nelson Fernández	Universidad de Pamplona
Emilio Ferrara	University of Southern California
Pedro Ferreira	Carnegie Mellon University
Alessandro Flammini	Indiana University Bloomington
Francesc Font-Clos	ISI Foundation
Koen Frenken	Utrecht University
Seth Frey	Dartmouth University
Javier Galeano	Universidad Politécnica de Madrid
Jianxi Gao	Rensselaer Polytechnic Institute
Oliva Garcia Cantu Ros	Nommon
Rodrigo García-Herrera	LANCIS
Simon Garnier	New Jersey Institute of Technology
Georgi Georgiev	Assumption College and Tufts University
Carlos Gershenson	Universidad Nacional Autónoma de México
Gourab Ghoshal	University of Rochester
William Glenney	US Naval War College
Jesus Gomez-Gardenes	Universidad de Zaragoza
Arnaud Grignard	Massachusetts Institute of Technology
Patrick Grim	SUNY Stony Brook
Todd Guilfoos	University of Rhode Island
Martin Hanczyc	University of Trento
Laurent Hébert-Dufresne	University of Vermont
Leila Hedayatifar	New England Complex Systems Institute

César Hidalgo	Massachusetts Institute of Technology
Hiroyuki Iizuka	Hokkaido University
Gerardo Iñiguez	Universidad Nacional Autónoma de México
Luis R. Izquierdo	University of Burgos
Segismundo Izquierdo	Universidad de Valladolid
Klaus Jaffe	Universidad Simón Bolívar
Wander Jager	University of Groningen
Marco Alberto Javarone	University of Kent
Hang-Hyun Jo	Asia Pacific Center for Theoretical Physics
Raja Kali	University of Arkansas
Márton Karsai	ENS de Lyon
Mikko Kivela	Aalto University
Mark Kon	Boston University
Matthijs Koopmans	Mercy College
Andreas Krause	University of Bath
Lesley Kuhn	University of Western Sydney
Marcelo Kuperman	Centro Atomico Bariloche
Renaud Lambiotte	University of Oxford
Bruno Lara	Universidad Autonoma del Estado Morelos
Felipe Lara-Rosano	Universidad Nacional Autonoma de Mexico
María Elena Lárraga	Universidad Nacional Autonoma de Mexico
Michael Lees	University of Amsterdam
Erika Fille Legara	Asian Institute of Management
Sune Lehmann	Technical University of Denmark
Stephan Leitner	Alpen-Adria Universitaet Klagenfurt
Tom Lenaerts	Universite Libre de Bruxelles
Bruno Lepri	FBK-Irst
San Liang	Nanjing Institute of Meteorology
Benyamin Lichtenstein	University of Massachusetts Boston
Antonio Lopolito	Università di Foggia
Vittorio Loreto	Sapienza Universita' di Roma
Jorge Louçã	ISCTE
Miguel Luengo-Oroz	UN Global Pulse
Pablo Marquet	Pontificia Universiad Catolica de Chile
Serafin Martinez Jaramillo	Banco de Mexico
Luis A. Martinez Vaquero	Institute of Cognitive Sciences and Technologies
Angels Massip-Bonet	Universitat de Barcelona
Rossana Mastrandrea	IMT Institute for Advanced Studies Lucca
Jose L. Mateos	Universidad Nacional Autónoma de México
Philippe Mathieu	University of Lille 1
Raúl Alejandro Mejía Pedroza	INMEGEN
Jose Fernando Mendes	University of Aveiro
Veena Mendiratta	Nokia Bell Labs
Ralf Metzler	University of Potsdam

Ali Minai	University of Cincinnati
José M. Miotto	Max Planck Institute for the Physics of Complex Systems
Yosune Miquelajauregui	Universidad Nacional Autónoma de México
Pamela Mischen	Binghamton University
Marija Mitrovic	Institute of Physics Belgrade
Saurabh Mittal	The MITRE Corporation
Alfredo Morales	New England Complex Systems Institute
Yamir Moreno	University of Zaragoza
Esteban Moro	Universidad Carlos III de Madrid
Adilson Motter	Northwestern University
Maryluz Mouronte-López	Universidad Francisco de Vitoria
Sarah Muldoon	University at Buffalo
Markus Franziskus Müller	Universidad Autónoma del Estado de Morelos
Roberto Murcio	University College London
Mirco Musolesi	University College London
Andrea Nanetti	Nanyang Technological University
Muaz Niazi	COMSATS Institute of IT
Michael Norman	The MITRE Corporation
Elisa Omodei	UNICEF
Andreas Pape	Binghamton University
David Pastor-Escuredo	Universidad Politécnica de Madrid
Matjaz Perc	University of Maribor
Juan Ignacio Perotti	IMT Institute for Advanced Studies Lucca
Nicola Perra	University of Greenwich
Giovanni Petri	ISI Foundation
Sabine Pfleger	Universidad Nacional Autónoma de México
Andrew Philippides	University of Sussex
Carlos Adolfo Piña García	Universidad Veracruzana
Flavio Pinheiro	Massachusetts Institute of Technology
Daniel Polani	University of Hertfordshire
Chiara Poletto	Sorbonne Université
Tobias Preis	University of Warwick
Francisco Prieto-Castrillo	Massachusetts Institute of Technology
Denise Pumain	University Paris I
Rick Quax	University of Amsterdam
Jose J. Ramasco	Universitat de les Illes Balears
William Rand	North Carolina State University
Fabio Revuelta	Universidad Politécnica de Madrid
Bernard Ricca	St. John Fisher College
Luca Riccetti	Università degli studi di Macerata
Alberto Robledo	Universidad Nacional Autónoma de México
Carlos Rodriguez-Sickert	Universidad del Desarrollo
Jonathan Roginski	West Point Network Science Center
Daniel Romero	University of Michigan

Andrea Roventini	Scuola Superiore Sant'Anna
Celine Rozenblat	University of Lausanne
Stefano Ruffo	SISSA
Alberto Russo	Università Politecnica delle Marche
Alex Rutherford	Masdar Institute of Science and Technology
Gonzalo Ruz	Universidad Adolfo Ibáñez
Hadrien Salat	University College London
Maxi San Miguel	Universitat de les Illes Balears
Marc Santolini	Northeastern University
Francisco C. Santos	Universidade de Lisboa
José I. Santos	University of Burgos
Jari Saramäki	Aalto University
Hiroki Sayama	Binghamton University
Samuel Scarpino	University of Vermont
Andreas Schadschneider	Universitaet zu Koeln
Jeff Schank	University of California
Frank Schweitzer	ETH Zurich
M. Ángeles Serrano	Universitat de Barcelona
Arturo Serrano-Santoyo	CICESE Research Center
Saray Shai	Wesleyan University
Yakov Shapiro	University of Alberta
Paul Silvey	The MITRE Corporation
J. Mario Siqueiros-García	Universidad Nacional Autónoma de México
Zbigniew Smoreda	Orange Labs
Stanislav Sobolevsky	New York University
Sorin Solomon	Hebrew University of Jerusalem
Pablo Suárez-Serrato	Universidad Nacional Autónoma de México
Pål Sundsøy	NBIM
Michael Szell	Central European University
Dane Taylor	University at Buffalo
Gabriele Tedeschi	Universita Politecnica delle Marche Ancona
Andrea Teglio	Universitat Jaume I
Michele Tizzoni	ISI Foundation
Andreas Tolk	The MITRE Organization
Emma Towlson	Northeastern University
Stephen Uzzo	New York Hall of Science
Sergi Valverde	University Pompeu Fabra
Gábor Vásárhelyi	Eötvös University
Nick Vriend	Queen Mary University of London
Hui Yang	Pennsylvania State University
Taha Yasseri	University of Oxford
Hyejin Youn	Kellogg School of Management
Burcu Yucesoy	Northeastern University

Stefano Zapperi	University of Milan
Haifeng Zhang	Carnegie Mellon University
Zi-Ang Zhang	Binghamton University
Chen Zhong	King's College London
Arkaitz Zubiaga	University of Warwick

Contents

Epistemological Constraints When Evaluating Ontological Emergence with Computational Complex Adaptive Systems

Andreas Tolk(✉), Matthew T. K. Koehler, and Michael D. Norman

The MITRE Corporation, Hampton, VA 23666, USA
atolk@mitre.org

Abstract. Natural complex adaptive systems are of particular scientific interest in many domains, as they may produce something new, like structures, patterns, or properties, that arise from the rules of self-organization. These novelties are emergent if they cannot be understood as any property of the components, but as a new property of the system. One of the leading methods to better understand complex adaptive systems is the use of their computational representation. In this paper, we make the case that emergence in computational complex adaptive systems can only be epistemological, as the constraints of computer functions do not allow for the creation of something new, as required for ontological emergence. As such, computer representations of complex adaptive systems are limited in producing emergence, but nonetheless useful to better understand the relationship between emergence and complex adaptive systems.

Keywords: Complexity · Epistemological emergence
Ontological emergence

1 Introduction

Complex adaptive systems are of interest to many scientists and researchers in many domains. Buckley was among the first using the term complex adaptive system [6]. He applied systems research methods to better understand behavior in social systems, as the often used linear, categorical descriptions of processes and interactions did not sufficiently explain the complex nature of the subject of discourse. Of particular interest to Buckley was the observation of emergence in such systems.

The recent developments of computational methods supporting scientific research have led to the rise of a variety of computational science disciplines, and in particular to the increased use of computational methods to evaluate

A. J. Morales et al. (Eds.): ICCS 2018, SPCOM, pp. 1–10, 2018.
https://doi.org/10.1007/978-3-319-96661-8_1

complex adaptive systems, as discussed in significant detail in [1,14,23]. In particular, agent-based modeling is often used to implement, computationally, complex adaptive systems; thus, allowing one to observe emergent macro-level system behavior that is not formulated explicitly, but rather results from the many micro-level interactions of the agents. As a result, agent-based models are often seen as the tool of choice for computational social science. As Bankes captures it in his introduction to the proceedings of the National Academy of Sciences on this topic:

> In social science, topics such as the emergence of cultural norms or institutions from the interaction of individual activity are indeed very important and not well addressed by competing modeling formalisms. So, the demonstrated ability of Agent Based Modeling to discover examples of such emergent dynamics from knowledge about the behavior of members of a society is potentially quite useful [2].

However, many researchers in these application domains do not have a formal computer science education and are not fully aware of many of the underlying principles from the philosophy of science. As a result, they do not only use simulation as a reference to study complex adaptive systems, rather they interpret observations of computationally instantiated complex adaptive systems to fully and equally represent their natural counterparts. In their perception, the simulation "replaces" the real system. Knowledge gained from the computational experiments is mapped directly to insights and applications of the real systems. Studying the computational representation becomes equivalent to studying the underlying system in the real world rather than as the generation of sufficiency theorem [1] related to the real world system.

Based on the research conducted in support of [32], this paper gives examples of natural and computational complex adaptive systems and observable emergence, introduces a critical review of the categories of emergence from the philosophy of science perspective, and concludes that there are *significant epistemological constraints* when computational methods are used for evaluating emergence.

2 Complex Adaptive Systems

The literature agrees that simple systems behave in a straightforward, mechanical, usually linear, and, most of all, easy to predict and manner: they behave as expected. A complicated system is composed of many often nonlinearly interacting parts that can be studied using reductionist and probabilistic models and statistical methods. It is still predictable, but it usually requires experts who are highly educated and experienced and have a tailored tool set available.

For the definition of complex systems, in [29] the authors propose the following definition after a review of systems engineering-relevant literature from complexity science:

> Complex systems are systems that do not have a centralizing authority and are not designed from a known specification, but instead involve disparate stakeholders creating systems that are functional for other purposes and are only brought together in the complex system because the individual "agents" of the system see such cooperation as being beneficial for them.

Complex systems are not easily predictable, and the principles of reductionism do not bear fruit when laboring to understand them, as system behavior emerges on all levels of the system. Although they are not fully knowable, within reason there may be some prediction possible.

Complex *adaptive* systems add the element of self-organized adjustment of some or all of its components and of the system itself. For its definition, the focus very often is the agent metaphor for the system components, as compiled in [9] and revisited in [5]. One of the insights drawn from these overview articles is that the diversity of definitions suggests one should focus on properties of such systems, as elaborated in detail within Holland's seminal contributions to the unified theory of complex adaptive systems [15], which contained a particular focus on aggregation – complexity emerges from the interaction of smaller components, which themselves may be the products of systems – and nonlinearity – agents interact in dynamical and non-linear ways.

The reason we use computational science in general, and in the context of this paper computational complex adaptive systems, is that they help us to understand natural systems. There are legions of examples of natural complex adaptive systems that have been evaluated using computational support. Without claiming completeness, some examples include, *inter alia:*

- society [6],
- the ecosystem and biosphere [19],
- supply networks [8],
- human language [3,33],
- product development environments [22],
- health care [28],
- climate change [17], and
- urban hazard mitigation [12].

All these systems are not fully knowable and often quite hard to predict. Moreover, it is often difficult to even collect useful data about these systems. Unfortunately, all of these systems (and many others) are very important, so we cannot ignore them. We must try to get a better understanding of their dynamics and causal structures, and we want to provide decision makers a better foundation with which to make informed decisions. Currently, the most powerful tool to do this is the use of computational representations of these systems and to simulate their dynamics [7].

As already discussed in the introduction, computational science disciplines, such as computational physics, computational biology, computational chemistry, computational social science, and many more, explore the use of computer models and simulation in direct support of their research. The discipline of complex systems research benefit significantly from these developments, as computers amplify our abilities to model, simulate, and evaluate the *computational representations* of the systems of interest [31].

The principle steps of such a computational study of complex adaptive systems was captured in detail by [14] and professionalized by many authors since then. The Santa Fe Institute and other similar organizations dedicate their work to the multidisciplinary study of the fundamental principles of complex adaptive systems, including physical, computational, biological, and social systems. These multidisciplinary scholars and students are experts in their fields and come together to use computational complex adaptive systems in support of their research, and the resulting studies are impressive, such as studies conducted to prevent collapse of tropical forests [13], or new insights into how evolution works [26], just to name a few.

These studies are generally understood to prove the enormous value of computational complex adaptive systems, in particular when it comes to emergent behavior, which is a characteristic property of complex systems: system behavior that does not depend on its individual parts, but on the multiple relations and interactions on all system levels. The next section will provide a short overview of the different forms of emergence from a philosophical, as well as, from a systems engineering perspective.

3 Categories of Emergence

We understand emergence as an unpredictable macro-level behavior that dynamically arises from the spatio-temporal and multilevel interactions between the parts down to the micro-level of the system. These interactions may be a constraint on various levels within the system. Natural systems are open systems exposing emergent behavior to the observer. This novel, irreducible, and unpredictable macro-behavior adds further complexity to the system when it causes itself changes at the micro-level, which then can result in new behavior emerging on the macro-level. Overall, the system may adapt to a new environment by developing new multi-level interactions and feedback loops [18].

Philosophy has dealt with the challenge of emergence for more than a century, starting with George Henry Lewes foundational work [20], long before any computational system existed to support them. Philosophers distinguish between epistemological and ontological emergence. Of the two, ontological emergence became a far more active research thread than did epistemological emergence [30].

3.1 Epistemological and Ontological Emergence

Epistemology is the theory of gaining knowledge, its methods, validity, and its scope. Knowledge is understood as the scientifically justified belief in something.

In the epistemological view, emergent properties and laws are systemic features of complex systems. This system is governed by true, law-like generalizations within a special science that is irreducible to fundamental physical theory for conceptual reasons. What is hidden to the researchers are these laws, resulting in the unpredictability of the emergence. As such, this view characterizes the concept of emergence in terms of limits of the human knowledge about the laws governing complex systems. The exposure of novel properties and behaviors is a characteristic of the system, but the unpredictability is a matter of knowledge.

The ontological view is quite different. Essentially, the ontological view of emergent properties is premised upon the idea that they are independent of the human knowledge. Instead they are novel, fundamental types of properties in and of themselves. Something new emerges that was not there before, and that cannot be explained by the components and their interactions and relations alone. The occurrence of emergent properties is not in any sense constituted by the occurrence of more fundamental properties and relations of the object's parts. Something really new is emerging from the system that has not been there before.

A detailed discussion on the philosophical views of emergent properties and their interpretation under epistemological and ontological viewpoints is presented in [25]. For the purposes of this paper, we will use a rather simple view, specifically: that epistemological emergence can be reduced by gaining more knowledge about the system, while ontological emergence cannot, because it is an inherit characteristic of the system.

3.2 Maier's Emergence Categories

This section introduces a systems engineering perspective. Mark Maier is known for his contributions to the discussion about systems of systems. In [21], he defined four categories of emergence, depending upon how well the emergence observed in the natural system can be reproduced and explained through a computational system.

- Simple emergence: The emergent property or behavior is predictable by simplified models of the system's components.
- Weak emergence: The emergent property is readily and consistently reproduced in simulation of the system but not in reduced complexity non-simulation models of the system, that is, simulation is necessary to reproduce it.
- Strong emergence: The emergent property is consistent with the known properties but, even in simulation, is inconsistently reproduced without any justification of its manifestation.
- Spooky emergence: The emergent property is inconsistent with the known properties of the system and is impossible to reproduce in a simulation of a model of equal complexity as the real system.

As implied by the terms used above, systems engineers normally do not like emergent phenomena as it results in behavior that is unforeseen and unpredictable. Consequently, Maier's viewpoint is written from the position that emergence is be avoided or, at least, controlled. In contrast, as engineers of complex adaptive systems, we seek to enable and leverage emergence [24].

In [32] we observe that simple emergence corresponds with congruent computable systems, and weak emergence with complicated computable system, where both system categories are predictable, at least in hindsight. Strong emergence falls into the category of unpredictable complex systems, while Maier's "spooky" emergence lies even outside of our system thinking boundaries, referencing real ontological emergence phenomena. Simple and weak emergence are epistemological, strong and spooky are ontological. The following figure captures these relations (Fig. 1).

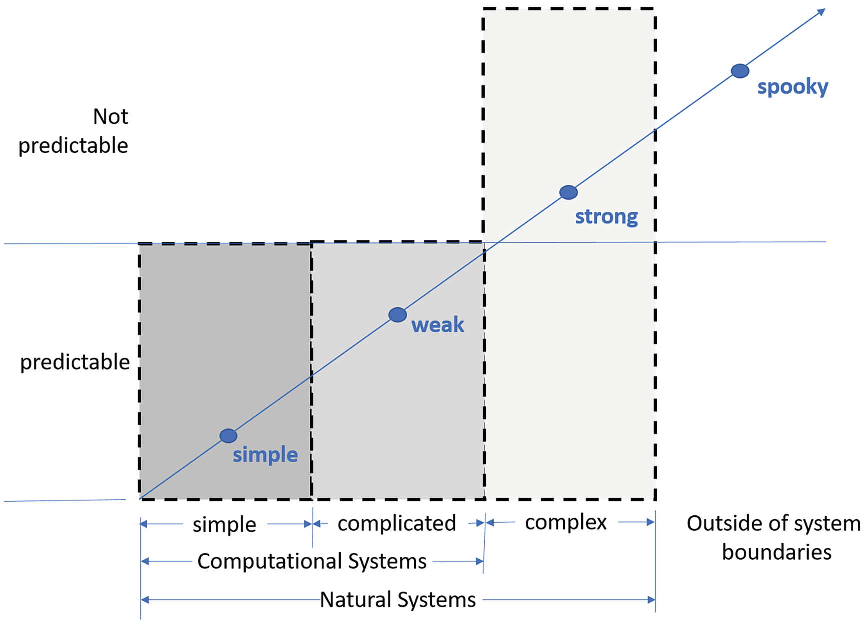


Fig. 1. Systems and emergence

4 Epistemological Constraints of Computational Systems

In the light of this discussion, it seems to be fair to look at how we can gain insight and knowledge from the application of computational systems, such as modeling and simulation, and, in particular, agent-based approaches.

The famous Artificial Intelligence researcher Huber L. Dreyfus is well known for his two books on the limits and constraints of computers: "What Computers Can't Do: The Limits of Artificial Intelligence" [10] and "What Computers Still Can't Do: A Critique of Artificial Reason" [11]. Dreyfus points to known limits that are founded in the nature of computers as captured in the works of Turing, Church, Gödel, and other pioneers of computer science that were often overlooked by his colleagues. These constraints are, of course, still valid for computational platforms in general despite the advances in hardware and software that have been made over the years. Furthermore, these constraints extend to systems represented within a computational platform, including representations of complex adaptive systems.

The essential constraint remains: computers transform input parameters into output parameters using computable functions to do so. As a result, computers cannot create something new out of nothing, as everything that is produced by a computer must be in the input data or the transforming algorithm. As Boden points out in her paper, we can model combinational, exploratory, and transformational forms of creativity [4]. However, all these forms creativity are discovering through rearrangement and transformation, not creating something new that was not there before.

As such, the question arises if computational systems can indeed be complex in the sense of producing emergence? They can be complicated, even extremely complicated, but in principle, it boils down to the transformation of input parameters into output parameters using computational (computable) functions. That is true for all computations, including agent-based models. These are insights covered by peer reviewed, often foundational publications and as such important for computational complex adaptive systems as well.

5 Discussion

These observations raise the question if computational complex adaptive systems cannot produce ontological emergence, are they still useful? As observed by Rouse in [27], the borders between the system categories - simple, complicated, and complex - are often fuzzy and depend on the education and experience of the team. What looks complex and unpredictable to a novice may turn out to be complicated at best for an expert team. This view is true without question for epistemological emergence: if we increase the human knowledge, we reduce the epistemological emergence, so the application of computational complex adaptive systems is very useful!

In the best case, we can close all knowledge gaps, resulting in a significant reduction of complexity, comparable to moving from an epicycle model of the solar system to a Copernican and Kepler model. If all we observe in natural complex adaptive systems can be explained using computational representations, that would be a tremendous accomplishment. Humphreys shows in [16] how we use computational means to extend our abilities to gain knowledge and produce new scientific insights. We just have to be careful to remember that this insight does not come by exactly reproducing natural complex adaptive systems, but they help to reduce epistemological emergence by increasing our knowledge.

Finally, as discussed in [32], complexity is often a multivariable and multidimensional phenomenon within the space-time continuum. Multivariable implies an often-large number of variables with sometimes incomplete knowledge about their interdependencies. Multidimensional implies possibly multiple vantage points that are dependent on the frame of reference when observing the phenomenon. These vantage points are not mutually exclusive, but rather focus different facets. Furthermore, the phenomenon may manifest over time or over space in the space-time continuum, requiring methods allowing for spatio-temporal analysis instead of exclusively looking at local snapshots.

Where does this leave us and our ability to study and understand complex systems? Given the aforementioned characteristics, when analyzing complex adaptive systems using computationally-based modeling and simulation methods are still the best options we have today. However, given the fundamental limits to computational systems, and indeed any formal system, it is likely that there are classes of complex systems that cannot be usefully represented in a computational (Turing machine) form.

While that may be true, in most cases complex systems and their emergent phenomena come from understood components and interact in knowable ways. This being the case, for most purposes complex systems can be meaningfully studied by computational methods. However, we must keep in mind that these are models, not the real system. What we create are "sufficiency theorems" [1]. We are deducing an outcome from an input and set of transformation rules, these rules may or may not be how nature actually works. Essentially, we can use computational methods to create deductions that can then be collected for inductive conclusions to solve abductive problems. But, just like all analytic methods, these are models of the real system in question. At the other extreme, one may argue if true ontological emergence in which something new emerges out of nothing is not magic, but rather some hidden law(s) we do not yet understand. As Arthur C. Clarke formulated this idea in his third law: *"Any sufficiently advanced technology is indistinguishable from magic."* Maybe computers can help us to advance the technology, but they will not produce ontological emergence.

Here we made the argument that, while there are fundamental limits to computation that puts a hard constraint on the complex systems that may be fruitfully represented within a computational system, many complex systems may be represented and studied with computational methods. However, we must not fall victim to the methods and remember that they are simply models that should be used to improve our understanding of complex systems. As all claims in this paper are based on accepted and peer reviewed - and often foundational - literature, additional experimental proofs are not necessary.

Disclaimer

The authors' affiliation with The MITRE Corporation is provided for identification purposes only, and is not intended to convey or imply MITRE's concurrence

with, or support for, the positions, opinions, or viewpoints expressed by the authors. This paper has been approved for Public Release; Distribution Unlimited; Case Number 17-3081-15.

References

1. Axtell, R.: Why Agents? on the Varied Motivations for Agent Computing in the Social Sciences. Center on Social and Economic Dynamics Brookings Institution, Washington, DC (2000)
2. Bankes, S.C.: Agent-based modeling: a revolution? Proc. Natl. Acad. Sci. **99**(suppl 3), 7199–7200 (2002)
3. Beckner, C., Blythe, R., Bybee, J., Christiansen, M.H., Croft, W., Ellis, N.C., Holland, J., Ke, J., Larsen-Freeman, D., Schoenemann, T.: Language is a complex adaptive system: position paper. Lang. Learn. **59**(s1), 1–26 (2009)
4. Boden, M.A.: Computer models of creativity. AI Mag. **30**(3), 23 (2009)
5. Brownlee, J.: Complex adaptive systems. Complex Intelligent Systems Laboratory, Centre for Information Technology Research, Faculty of Information Communication Technology, Swinburne University of Technology, Melbourne, Australia (2007)
6. Buckley, W.: Society as a complex adaptive system. In: Modern Systems Research for the Behavioral Scientist. Aldine Publishing Company (1968)
7. Buss, S., Papadimitriou, C.H., Tsitsiklis, N.: On the predictability of coupled automata: an allegory about chaos. In: 1990 Proceedings of the 31st Annual Symposium on Foundations of Computer Science, pp. 788–793. IEEE (1990)
8. Choi, T.Y., Dooley, K.J., Rungtusanatham, M.: Supply networks and complex adaptive systems: control versus emergence. J. Oper. Manage. **19**(3), 351–366 (2001)
9. Dooley, K.: Complex adaptive systems: a nominal definition. Chaos Netw. **8**(1), 2–3 (1996)
10. Dreyfus, H.L.: What Computers Can't Do: The Limits of Artificial Intelligence. Harper & Row, New York (1972)
11. Dreyfus, H.L.: What Computers Still Can't Do: A Critique of Artificial Reason. MIT Press, Boston (1992)
12. Godschalk, D.R.: Urban hazard mitigation: creating resilient cities. Nat. Hazards Rev. **4**(3), 136–143 (2003)
13. Hébert-Dufresne, L., Pellegrini, A.F., Bhat, U., Redner, S., Pacala, S.W., Berdahl, A.M.: Edge fires drive the shape and stability of tropical forests. Ecol. Lett. **21**(6), 794–803 (2018)
14. Holland, J.H.: Complex adaptive systems. Daedalus 17–30 (1992)
15. Holland, J.H.: Hidden Order: How Adaptation Builds Complexity. Addison Wesley Publishing Company, Boston (1995)
16. Humphreys, P.: Extending Ourselves: Computational Science, Empiricism, and Scientific Method. Oxford University Press, Oxford (2004)
17. Ingwersen, W.W., Garmestani, A.S., Gonzalez, M.A., Templeton, J.J.: A systems perspective on responses to climate change. Clean Technol. Environ. Policy **16**(4), 719–730 (2014)
18. Kauffman, S.A.: The origins of order: self-organization and selection in evolution. In: Spin Glasses and Biology, pp. 61–100. World Scientific (1992)
19. Levin, S.A.: Ecosystems and the biosphere as complex adaptive systems. Ecosystems **1**(5), 431–436 (1998)

20. Lewes, G.H.: Problems of Life and Mind. Trübner & Company, London (1877)
21. Maier, M.W.: The role of modeling and simulation in system of systems development. In: Rainey, L.B., Tolk, A. (eds.) Modeling and Simulation Support for System of Systems Engineering Applications. Wiley (2015)
22. McCarthy, I.P., Tsinopoulos, C., Allen, P., Rose-Anderssen, C.: New product development as a complex adaptive system of decisions. J. Prod. Innov. Manage. **23**(5), 437–456 (2006)
23. Miller, J.H., Page, S.E.: Complex Adaptive Systems: An Introduction to Computational Models of Social Life: An Introduction To Computational Models of Social Life. Princeton University Press, Princeton (2009)
24. Norman, M.D., Koehler, M.T., Pitsko, R.: Applied complexity science: enabling emergence through heuristics and simulations. In: Emergent Behavior in Complex Systems Engineering: A Modeling and Simulation Approach, pp. 201–226 (2018)
25. O'Connor, T., Wong, H.Y.: Emergent properties. In: Zalta, E.N. (ed.) The Stanford Encyclopedia of Philosophy, Summer 2015 edn. Metaphysics Research Lab, Stanford University (2015)
26. Payne, J.L., Khalid, F., Wagner, A.: RNA-mediated gene regulation is less evolvable than transcriptional regulation. Proc. Natl. Acad. Sci. 201719138 (2018)
27. Rouse, W.B.: Engineering complex systems: implications for research in systems engineering. IEEE Trans. Syst. Man Cybern. Part C (Appl. Rev.) **33**(2), 154–156 (2003)
28. Rouse, W.B.: Health care as a complex adaptive system: implications for design and management. Bridge Wash. Natl. Acad. Eng. **38**(1), 17 (2008)
29. Sheard, S.A., Mostashari, A.: Principles of complex systems for systems engineering. Syst. Eng. **12**(4), 295–311 (2009)
30. Silberstein, M., McGeever, J.: The search for ontological emergence. Philos. Q. **49**(195), 201–214 (1999)
31. Tolk, A.: Simulation and modeling as the essence of computational science. In: Proceedings of the 50th Summer Computer Simulation Conference (2018)
32. Tolk, A., Diallo, S., Mittal, S.: Complex systems engineering and the challenge of emergence. In: Emergent Behavior in Complex Systems Engineering: A Modeling and Simulation Approach, pp. 79–97. Wiley (2018)
33. Zeigler, B.P., Mittal, S.: System theoretic foundations for emerging behavior modeling: the case of emergence of human language in a resource-constrained complex intelligent dynamical system. In: Emergent Behavior in Complex Systems Engineering: A Modeling and Simulation Approach, pp. 35–57. Wiley (2018)

Synergy and the Bioeconomics of Complexity

Peter A. Corning(✉)

The Institute for the Study of Complex Systems
is headquartered in Seattle, Washington, USA
pacorning@complexsystems.org

Abstract. Living systems are distinctive in that they are subject to basic economic criteria, and to economic constraints. They are obedient to the calculus of economic costs and benefits in any given environmental context. This applies to all biological traits, including complexity (which can be defined and measured in both structural and functional terms). A major theoretical challenge, therefore, is to account for the "progressive" evolution of complex living systems over time, from the origins of life itself to "superorganisms" like leaf cutter ants and humankind. Why has complexity evolved? A causal theory, called the Synergism Hypothesis, was first proposed by this author in the 1980s and was independently proposed by John Maynard Smith and Eörs Szathmáry in the 1990s. This theory is only now emerging from the shadows as a major paradigm shift is occurring in evolutionary biology away from a reductionist, individualistic, gene-centered model to a multi-level, systems perspective. The Synergism Hypothesis is, in effect, an economic (or bioeconomic) theory of complexity. It is focused on the costs and benefits of complexity, and the unique creative power of functional synergy in the natural world. The theory proposes that the overall trajectory of the evolutionary process over the past 3.8 billion years or so has been shaped by synergies of various kinds. The synergies produced by cooperation among various elements, genes, parts, or individuals may create interdependent "units" of adaptation and evolutionary change that are favored in a dynamic that Maynard Smith termed Synergistic Selection (in effect, a sub-category of natural selection). Some methodological issues will also be discussed, and some examples will be provided.

1 Introduction

Much of the work in complexity science in recent years has been focused on the physical, structural, functional, and dynamical aspects of complex phenomena, as reflected in the papers for this volume. Well and good. Living systems are, after all, embedded in the physical world.

However, complex organisms are distinctive in that they are also subject to basic economic criteria, and to economic constraints. Biological complexity is not simply an end in itself, nor an historical artifact, much less the product of some exogenous physical trend, force, or "law". Over the years, many candidate laws have been proposed that have claimed to explain complexity in evolution, going back to Jean Baptiste de Lamarck's "power of life" and Herbert Spencer's "universal law of evolution" in the nineteenth century (see the discussion in Corning 2018). In the latter part

A. J. Morales et al. (Eds.): ICCS 2018, SPCOM, pp. 11–23, 2018.
https://doi.org/10.1007/978-3-319-96661-8_2

of the twentieth century, the development of new mathematical tools and rise of complexity theory in various disciplines inspired a plethora of new law-like, or mechanistic explanations. This theme has continued into the new century. Perhaps most provocative is physiologist John Torday's (2016) claim that biological complexity is only an "epiphenomenon" of a mechanistic dynamic associated with physiological homeostasis (see also Torday and Rehan 2017).[1]

The problem with all such deterministic theories is that they explain away the very thing that needs to be explained – namely, the contingent nature of living systems and their fundamentally functional, adaptive properties. The purveyors of these theories often seem oblivious to the inescapable challenges associated with what Darwin called the "struggle for existence" in the natural world, and they discount the economics – the costs and benefits of complexity. Nor can they explain the fact that some 99% of all the species that have ever evolved are now extinct. Life is phenomenon that is at all times subject to the requirement that the bioeconomic benefits (direct or indirect) of any character or trait – including complexity – must outweigh the costs. It is subject to

[1] Though he is antagonistic to traditional neo-Darwinism, Torday's theory tacitly acknowledges the role of differential selection (natural selection) in evolution. His scenario posits the development of a set of highly synergistic physiological components that combined to produce homeostasis, which, he claims, has driven further physiological developments over time. But phrases like "selection advantage" and "positive selection" sneak into his discussion at various points. For example, he describes the over-engineering of lung capacity in land animals as being, very likely, the result of evolution "positively selecting for those organisms with optimal exchange capacity." It's natural selection in deep disguise.

A comment is also in order here regarding the mechanical engineer Bejan's (2016) much-hyped new theory of "everything" (his term) in physics, which he calls the "constructal law of design in nature." Bejan's claim is that there is a universal, inherent tendency for any "flow system" in nature – from rivers to living systems – to evolve over time in such a way as to provide "ever greater access to the currents that flow through it." Take, for example, the energy throughputs in living organisms. To Bejan, the increases in energy flows over the course of biological evolution accord with his physical law. It resembles similar physical trends. To a biologist, however, any increases in energy flows over time have had a strictly functional basis. An increase in efficiency, or in energy throughputs, is the end-product of natural selection – differential survival and reproduction among naturally occurring variations in energy capture/utilization capacities. Bejan's theory only accounts for the "winners". But, in reality, this is an artifact of the functional advantages involved and not of some exogenous "law". Indeed, Bejan's "flow" model cannot predict major functional variations in living systems. Consider water consumption. Filter feeders, like sponges, can process huge quantities of water in the course of a day, typically more than their own weight every few seconds. A human consumes only a small fraction of that amount, by weight. Moreover, the actual water throughput for any individual human is very much context-dependent. A marathon runner on a hot day will consume much more water (perhaps two quarts per hour) than a sedentary person of the same weight who is watching TV in an air-conditioned living room. It is the same with energy throughputs. Indeed, various specialized cells in our bodies consume vastly different amounts of energy (see below).

functional criteria and the calculus of economic costs and benefits in any given environmental context.[2]

So, the question is, what are the advantages of biological complexity? One (flippant) answer is that, if you assemble just the right "package" of attributes, you can create a human. A proper answer is much more complicated – of course. We need to start by defining what complexity means in relation to living systems and then examine how – and why – biological complexity has evolved over time.

2 Defining Biological Complexity

The question of how to define biological complexity has been much-debated over the years. It is evident that there is no one correct way to measure it; it can be defined in different ways for different purposes. However, two alternative methodologies are relevant (at least in theory) as ways of characterizing the broad evolutionary trend toward multi-leveled complex systems over the past 3.8 billion years or so, beginning with the origins of life and culminating (temporally at least) in humankind.

One method is structural. A synthetic complexity scale can be constructed from the number of levels of organization (inclusive of social organization), the number of distinct "parts", the number of different kinds of parts, and the number of interconnections among the parts (see Corning and Szathmáry 2015). The other method is functional. A complexity scale can be derived from the number of functionally discrete "tasks" in the division/combination of labor at all levels of organization, coupled with the quantity of "control information" that is generated and utilized by the system. Control information is defined as "the capacity to control the capacity to do work" in a cybernetic process; it is equivalent to the amount of thermodynamic work that a system can perform (see Corning and Kline 1998; Corning 2005, 2007). Both of these methodologies are relevant for the theoretical paradigm that will be discussed here.

3 Measuring the Costs and Benefits

There are also various ways of measuring the economic costs and benefits of biological complexity. The "ultimate" measure of profitability is, of course, reproductive success. Although the level of personal investment can vary widely in the natural world, an organism must sustain a minimal economic "profit" in order to be able to reproduce

[2] For example, Torday (2016) affirms that economic criteria have been operative even in the basic physiological evolution of living systems. One illustration: "…[I]t has been observed that the genome decreased by about 80%–90% after the Cambrian Extinction. The advent of endothermy may explain this phenomenon because ectotherms require complex enzymatic regulatory mechanisms in order to accommodate variable atmospheric temperatures, whereas the uniform body temperature of endotherms/homeotherms only requires one metabolic isoform to function optimally. Since metabolic genes account for 17% of the human genome, representing a fraction of the number of metabolic genes expressed by ectotherms, this reduction in metabolic enzyme heterogeneity would have contributed to the dramatic decrease in post-Cambrian genomic size." In other words, natural selection favored functional efficiencies/economies.

itself, and the more offspring it produces the more profitable it is from an ultimate evolutionary perspective.

However, there are also a many other, "proximate" ways of measuring the costs and benefits involved in "earning a living" in nature, and a number of familiar economic criteria are likely to have been important from a very early stage in the history of life on Earth — capital costs, amortization, operating costs and, most especially, strict economic profitability. The returns had to outweigh the costs. There is, of course, a large research literature in behavioral ecology and bioeconomics that is focused on just such proximate issues (see especially Davies, Krebs and West 2012, as well as such journals as *Behavioral Ecology, Behavioral Ecology and Sociobiology*, and the *Journal of Bioeconomics*).

Consider the fundamental need for energy capture. Dating back to Erwin Schrödinger's classic lectures and small book, *What is Life?* in 1944, it has long been appreciated that thermodynamics is of central importance in understanding the nature of life, and the challenges of living. Living systems must do work and are subject to thermodynamic entropy and the Second Law. This imposes significant functional requirements.

However, there is also a deep tradition in biophysics that assumes away the economic challenges involved in creating "negative entropy" (Schrödinger's neologism for how living systems contradict the Second Law). Indeed, there is a school of theorists who have advanced the proposition that energy is somehow a free good and that available energy itself "drives" the process of creating order and organization in the living world (see for examples, Morowitz 1968; Kauffmann 1995; Holland 1998; Schneider and Sagan 2005).

A famous experiment in physics, Maxwell's Demon, unwittingly demonstrated why this assumption is incorrect (see the detailed critique in Corning 2005, Chap. 13). In a nutshell, there is no way the Demon could create thermodynamic order "without the expenditure of work" (to use Maxwell's own, ill-considered claim for the Demon). Living systems must adhere to the first and only law (so far) of "thermoeconomics", namely, that the energetic benefits (the energy made available to the system to do work) must outweigh the costs required for capturing and utilizing it. From the very origins of life, energy has never been a free good (although initially the costs may have been exogenous to the system, or "externalities" – see Corning 2018). As biological complexity has increased over time, the work required to obtain and use energy to sustain the system has increased correspondingly (see the review in Corning 2005). Indeed, improvements in bioenergetic technologies represent a major theme in evolutionary history and in every case involved synergistic phenomena.

4 The Synergism Hypothesis

How, then, do we account for the evolution of biological complexity? Over the course of the past two decades, the subject of complexity has emerged as a major theme within mainstream evolutionary biology, and a search has been underway for "a Grand Unified Theory" – as biologist McShea (2015) characterizes it – that is consistent with Darwin's great vision.

As it happens, such a theory already exists. It was first proposed in *The Synergism Hypothesis: A Theory of Progressive Evolution* in 1983, and it involves an economic (or perhaps bioeconomic) theory of complexity. The same idea was later independently proposed in their two books on the "major transitions" in evolution. Simply stated, cooperative interactions of various kinds, however they may occur, can produce novel combined effects – *synergies* – with functional advantages that may, in turn, become direct causes of natural selection. The focus of the Synergism Hypothesis is on the favorable selection of synergistic "wholes" and the combinations of genes that produce these wholes. The parts (and their genes) that create these synergies may, in effect, become interdependent units of evolutionary change.

In other words, the Synergism Hypothesis is a theory about the unique combined effects produced by the relationships and interactions between things. I refer to it as Holistic Darwinism because it is entirely consistent with natural selection theory, *properly understood* (see the book-length elaboration in Corning 2005). Accordingly, it is the functional (economic) benefits associated with various kinds of synergistic effects in any given context that are the underlying cause of cooperative relationships – and of complex organization – in the natural world. The synergy produced by the whole provides the proximate functional payoffs that may differentially favor the survival and reproduction of the parts (and their genes).

Illustrates this idea with an analogy. The recipe for a biscuit/cookie is rather like the genome in living organisms. It represents a set of instructions for how to make an end-product. A shopper who buys a biscuit/cookie selects the "phenotype" – the end-product, not the recipe. So, if the recipe survives and the number of cookies multiply over time, it's only because shoppers like the end-product and are willing to purchase more of them. Although it may seem like backwards logic, the thesis is that functional synergy is the cause of cooperation and complexity in living systems, not the other way around.

5 Synergistic Selection

Maynard Smith also proposed the concept of Synergistic Selection in a 1982 paper as (in effect) a sub-category of natural selection. Synergistic Selection refers to the many contexts in nature where two or more genes/genomes/parts/individuals have a shared fate; they are functionally interdependent. Maynard Smith illustrated with a formal mathematical model that included a term for "non-additive" benefits. As I argue in my 2018 book, Synergistic Selection is an evolutionary dynamic with much wider scope even than Maynard Smith envisioned. It includes, among other things, many additive phenomena with combined threshold effects and, more important, many "qualitative novelties" that cannot even be expressed in quantitative terms. Synergistic Selection focuses our attention on the causal dynamics and selective outcomes when synergistic phenomena of various kinds arise in the natural world. For it is synergy, and Synergistic Selection, that has driven the evolution of cooperation and complexity in living systems over time, including especially the major transitions in evolution.

One example (among the many cited in my book) is the evolution of eukaryotes. Increased size and complexity can have many functional advantages in the natural world (see below), and eukaryotic cells, inclusive of their complex internal architecture, average some 10–15,000 times larger than the typical prokaryote. However, this huge size difference requires many orders of magnitude more energy, and the key to solving this functional imperative was a symbiotic (synergistic) union between an ancestral prokaryote and an ancestor of the specialized, energy producing mitochondria in modern eukaryotic cells. Not only was this novel combination of labor mutually beneficial for each of the two partners, but it created a pathway for expanding and multiplying those benefits many times over. Some specialized cells in complex organisms like humans may contain hundreds, or even thousands, of mitochondria. Liver cells, for instance, have some 2,500 mitochondria and muscle cells may have several times that number. I refer to it as a "synergy of scale." (See also Lane 2017. For related work on the evolution of multicellularity, see Ratcliff et al. 2012, 2015.)

6 The Creative Role of Synergy

It should be emphasized that many things can influence the likelihood of cooperation and synergy in the natural world – the ecological context, specific opportunities, competitive pressures, the risks (and costs) of cheating or parasitism, effective policing, genetic relatedness, biological "pre-adaptations", and especially the distribution of costs and benefits. However, an essential requisite for cooperation (and complexity) – is functional synergy. Just as natural selection is agnostic about the sources of the functional variations that can influence differential survival and reproduction, so the Synergism Hypothesis is agnostic about how synergistic effects can arise in nature. They could be self-organized; they could be a product of some chance variation; they could arise from a happenstance symbiotic relationship; or they could be the result of a purpose-driven behavioral innovation by some living organism.

It is also important to stress that there are many different kinds of synergy in the natural world, including (as noted above) synergies of scale (when larger numbers provide an otherwise unattainable collective advantage), threshold effects, functional complementarities, augmentation or facilitation (as with catalysts), joint environmental conditioning, risk- and cost-sharing, information-sharing, collective intelligence, animal-tool "symbiosis" and, of course, the many examples of a division of labor (or more accurately, a combination of labor) in the natural world. Indeed, many different synergies may be bundled together (a synergy of synergies) in a complex socially organized "superorganism" like leaf cutter ants or *Homo sapiens* (for details, see Corning 2018).

7 Quantifying Synergy

Synergistic effects can also be measured and quantified in various ways. In the biological world, they are predominantly related to survival and reproduction. Thus, hunting or foraging collaboratively – a behavior found in many insects, birds, fish and mammals – may increase the size of the prey that can be pursued, the likelihood of success in capturing prey or the collective probability of finding a "food patch." Collective action against potential predators – herding, communal nesting, synchronized reproduction, alarm calling, coordinated defensive measures, and more – may greatly reduce an individual animal's risk of becoming a meal for some other creature.

Likewise, shared defense of food resources – a practice common among social insects, birds, and social carnivores alike – may provide greater food security for all. Cooperation in nest-building, and in the nurturing and protection of the young, may significantly improve the collective odds of reproductive success. Coordinated movement and migration, including the use of formations to increase aerodynamic or hydrodynamic efficiency, may reduce individual energy expenditures and/or aid in navigation. Forming a coalition against competitors may improve the chances of acquiring a mate, or a nest-site, or access to needed resources (such as a watering-hole, a food patch, or potential prey). In all of these situations, it is the synergies that are responsible for achieving greater efficiencies and enhancing profitability.

8 Testing for Synergy

There are also various ways of testing for synergy. One method involves experiments, or "thought experiments" in which a major part is removed from the whole. In many cases (not all), a single deletion, subtraction or omission will be sufficient to eliminate the synergy. Take away the heme group from a hemoglobin molecule, or the mitochondria from a eukaryotic cell, or the all-important choanocytes from sponges, or, for that matter, remove a wheel from an automobile. The synergies will vanish.

Another method of testing for synergy derives from the fact that many adaptations, including those that are synergistic, are contingent and context specific, and that virtually all adaptations incur costs as well as benefits. To repeat, the benefits of any trait must, on balance, outweigh the costs; it must be profitable in terms of its impact on survival and reproduction. Thus, it may not make sense to form a herd, or a shoal, or a communal nest if there are no threatening predators in the neighborhood, especially if proximity encourages the spread of parasites or concentrates the competition for scarce resources. Nor does it make sense for emperor penguins in the Antarctic to huddle together for warmth at high-noon during the warm summer months, or for Mexican desert spiders to huddle against the threat of dehydration during the wet rainy season. And hunting as a group may not be advantageous if the prey is small and easily caught by an individual hunter without assistance.

Another way of testing for synergy involves the use of a standard research methodology in the life sciences and behavioral sciences alike – comparative studies. Often a direct comparison will allow for the precise measurement of a synergistic effect. Some of the many documented examples in the research literature include flatworms that can collectively detoxify a silver colloid solution that would otherwise be fatal to any individual alone; nest construction efficiencies that can be achieved by social wasps compared to individuals; lower predation rates in larger meerkat groups with more sentinels; higher pup survival rates in social groups of sea lions compared to isolated mating pairs; the hunting success of cooperating hyenas in contrast with those that fail to cooperate; the productivity of choanocytes in sponges compared to their very similar, free-swimming relatives called choanoflagellates, and the comparison between lichen partnerships and their independently-living cousins. Some of the most relevant examples can be found in comparative genomics (e.g., see Berens et al. 2015).

9 "Why Size Matters"

In his important book, *Why Size Matters,* Bonner (2006) focused on the critical role of size in evolution and, equally important, the close linkage between size and biological complexity as he defined it, namely, an internal cellular division of labor. Bonner's thesis was that increased complexity (thus defined) in living systems is driven by increases in size. "There are universal rules imposed by size," he tells us (p.x). He also asserts that "size is the supreme regulator of all matters biological" (p. 2). Indeed, "size is a prime mover in evolution…increased size requires changes in structure and function" (ibid.).

It is certainly true that there is an interplay between the physics of size and the engineering and functional challenges associated with building and maintaining a larger organism. Gravity is an obvious problem. And so is the problem of producing and diffusing greater quantities of oxygen, energy and nutrients throughout a much larger system. However, like many other monolithic theories, the truth in this case probably lies somewhere in the middle. Increased functional capabilities and efficiencies are also necessary as prerequisites for increased size, and the question of which came first might be resolved by viewing the causal dynamics from a longitudinal perspective – as a process of *reciprocal causation* over time (see especially Laland 2011, 2013). It is an argument that goes back to Darwin himself in *The Origin of Species.*

But more important, Bonner's hypothesis begs the question. Why have organisms grown larger over time? Why do we see a progression in evolutionary history from microscopic prokaryotes with their relatively simple internal division of labor to much larger, intricately organized and far more complex eukaryotes, then to multicellular organisms, and, finally, to organized societies composed of many individual organisms, sometimes numbering in the millions? The answer, in brief, is that size is not an end in itself. It arises because it confers various functional advantages – various synergies of scale. These may include such things as improved mobility, more effective food acquisition, more efficient and effective reproduction, and, not least, protection from predators.

10 A Classic Example

Consider, for example, the volvocines, a primitive order of aquatic green algae that form into tight-knit colonies resembling integrated organisms. One of the smallest of these colonies (*Gonium*) has only a handful of cells arranged in a disk, while the *Volvox* that give the volvocine line its name may have some 50–60,000 cells arranged in the shape of a hollow sphere that is visible to the naked eye. Each *Volvox* cell is independent, yet the colony-members collaborate closely. For instance, the entire colony is propelled by a thick outer coat of flagella that coordinate their exertions to keep the sphere moving and slowly spinning in the water – in other words, a synergy of scale.

Some of the synergies in the *Volvox* were documented in a study many years ago by Bell (1985), and in more recent studies by Michod (1999, 2007, 2011). The largest of the *Volvox* colonies have a division of labor between a multi-cellular body and segregated reproductive cells. Bell's analyses suggested some of the benefits. A division of labor and specialization facilitates growth, resulting in a much larger overall size. It also results in more efficient reproductive machinery (namely, a larger number of smaller germ cells). The large hollow enclosure in *Volvox* also allows a colony to provide a protective envelope for its daughter colonies; the offspring disperse only when the parental colony finally bursts apart.

But there is one other vitally important synergy of scale in *Volvox*. It turns out that their larger overall size results in a much greater survival rate than in the smaller *Gonium*. These algae are subject to predation from filter feeders like the ubiquitous copepods, but there is an upper limit to the prey size that their predators can consume. The larger, integrated, multi-cellular *Volvox* colonies are virtually immune to predation from the filter feeders.

11 Toward a Post-modern Evolutionary Synthesis

Many theorists these days are calling for a new post-modern, post-neo-Darwinian synthesis. Some advocate the adoption of a more elaborate "multilevel selection" model. Others speak of an "Extended Evolutionary Synthesis" that would include developmental processes and Lamarckian inheritance mechanisms, among other things (see Pigliucci and Müller 2010; Jablonka 2013). Noble (2013) has proposed what he calls an "Integrative Synthesis" that would include the role of physiology in the causal matrix.

Whatever the label, it is clear that a much more inclusive framework is needed, one that captures the full dynamics, and the interactions, among the many different causal influences at work in the natural world. We also need to view the evolutionary process in terms of multi-leveled systems – functional organizations of matter, energy, and information, from genomes to ecosystems. And we must recognize that the level of selection – of differential survival and reproduction – in this hierarchy of system levels is determined in each instance by a synergistic configuration, or network of causes. Indeed, the outcome in any given context may be a kind of vector sum of the causal forces that are at work at several different levels at once.

In the heyday of the Modern Synthesis in the twentieth century, the explanatory framework in evolutionary biology was often truncated to focus on genetic mutations, sexual recombination, and the mathematics of differential selection (changes in gene frequencies) in an interbreeding population. This mathematical framework, albeit with many refinements, remains the theoretical backbone of the discipline to this day. The fundamental problem is that it explains very little. Natural selection (properly understood) is not an external causal agency or a "mechanism". It is a metaphor – an umbrella term for a wide-open framework that encompasses whatever specific factors may influence biological continuity and change in any given environment. Equally important, it is no longer tenable to view genetic mutations as the primary source of creativity in evolution. There are many different sources of innovation. In the words of Noble 2014) and his co-authors: "DNA does not have a privileged place in the chain of causality."[3] (See also Woese 2004.)

12 An Inclusive Synthesis

What is needed going forward is a broadly ecumenical paradigm that would provide more of a work plan than a finished product. Perhaps it could be characterized as an Inclusive Synthesis. It would be an open-ended framework for explaining how,

[3] Over the past few decades the fundamental tenets of neo-Darwinism have been convincingly challenged. It seems that organisms are active participants in shaping the evolutionary process. There is now a paradigm shift under way from an atomistic, reductionist, gene-oriented, mechanistic (robotic) model to a systems perspective in which "purposeful" actions and informational processes are recognized as fundamental properties of living organisms at all levels. In his important book, *Evolution: A View from the 21st Century,* the leading microbiologist Shapiro (2011, 2009) argues that cells must be viewed as complex systems that control their own growth, reproduction and even shape their own evolution over time. He refers to it as a "systems engineering" perspective. Indeed, there is no discreet DNA unit that fits the neo-Darwinian model of a one-way, deterministic gene. Instead, the DNA in a cell represents a two-way, "read-write system" wherein various "coding sequences" are mobilized, aggregated, manipulated and even modified by other genomic control and regulatory molecules in ways that can influence the course of evolution itself. "We need to develop a new lexicon of terms based on a view of the cell as an active, sentient entity," Shapiro stresses. Echoing the views of a number of other theorists recently, he calls for "a deep rethinking of basic evolutionary concepts." Indeed, Shapiro cites some 32 different examples of what he refers to as "natural genetic engineering," including immune system responses, chromosomal rearrangements, diversity generating retroelements, the actions of mobile genetic elements called transposons, genome restructuring, whole genome duplication, and symbiotic DNA integration. As Shapiro emphasizes, ``The capacity of living organisms to alter their own heredity is undeniable. Our current ideas about evolution have to incorporate this basic fact of life.'

The well-known senior physiologist Noble (2012, 2013), in a recent paper, argues that all the basic assumptions underlying the Modern Synthesis and neo-Darwinism have been proven wrong. Specifically, (1) genetic changes are often very far from random and in many cases are directed by "epigenetic" (developmental) and environmental influences; (2) genetic changes are often not gradual and incremental (Noble cites, among other things, the radical effects of DNA transposons, which have been found in more than two-thirds of the human genome); (3) an accumulation of evidence for a Lamarckian inheritance of epigenetic influences that has now reached the flood stage; and (4) natural selection, rather than being gene focused, is in fact a complex multi-leveled process with many different levels and categories of causation. Woese and Goldenfeld (2009) in their critique of the modern synthesis characterize life as a "collective phenomenon." And evolutionary theorist Eva Jablonka and her colleagues (Jablonka et al. 1998; Jablonka and Raz 2009; Jablonka and Lamb 2014) identify four distinct "Lamarckian" modes of inheritance: (1) directed adaptive mutations, (2) the inheritance of characters acquired during development and the lifetime of the individual, (3) behavioral inheritance through social learning, and (4) language-based information transmission. It could be called the extended genome. In a recent review of the mounting evidence for this Lamarckian view, Jablonka (2013) concludes: "The existing knowledge of epigenetic systems leaves little doubt that non-genetic information can be transmitted through the germ line to the next generation, and that internal and external conditions influence what is transmitted and for how long." The developmental biologist West-Eberhard (2003) goes even further: "Genes are followers, not leaders, in adaptive evolution."

precisely, natural selection "does its work" in any given context (what causal factors influence adaptive changes). It would also represent an ongoing work-in-progress rather than a completed theoretical edifice. Nor would it seek to reduce natural selection ultimately to some simple formula or "mechanism". No single discipline (or model) can capture such a complex, multi-faceted narrative. In the longer run, our theoretical enterprise will require a synthesis and integration of the many different specialized and rapidly growing areas of knowledge (see Love 2010).

In the meantime, the historical process through which these multilevel biological systems have evolved over time can be framed as a sequence of major transitions in complexity – from the very origins of life itself to the emerging global society that humankind is now engaged in creating (for better or worse). And, at every level in this hierarchy, we can see the driving influence of synergy and Synergistic Selection. The arc of evolution bends toward synergy.

References

Bejan, A.: The Physics of Life: The Evolution of Everything. St. Martin's Press, New York (2016)

Bell, G.: Origin and early evolution of germ cells as illustrated by the volvocales. In: Halverson, H.O., Monroy, A. (eds.) Origin and Evolution of Sex, pp. 221–256. Alan R. Liss, New York (1985)

Berens, A.J., Hunt, J.H., Toth, A.L.: Comparative transcriptomics of convergent evolution: different genes but conserved pathways underlie caste phenotypes across lineages of eusocial insects. Mol. Biol. Evol. **32**(3), 690–703 (2015). https://doi.org/10.1093/molbev/msu330. Epub 2014 Dec 9

Bonner, J.T.: Why Size Matters: From Bacteria to Blue Whales. Princeton University Press, Princeton (2006)

Corning, P.A.: The Synergism Hypothesis: A Theory of Progressive Evolution. McGraw-Hill, New York (1983)

Corning, P.A.: Holistic Darwinism: Synergy, Cybernetics and the Bioeconomics of Evolution. University of Chicago Press, Chicago (2005)

Corning, P.A.: Control information theory: the 'Missing Link' in the science of cybernetics. Syst. Res. Behav. Sci. **24**, 297–311 (2007)

Corning, P.A.: Synergistic Selection: How Cooperation Has Shaped Evolution and the Rise of Humankind. World Scientific, Singapore (2018)

Corning, P.A., Szathmáry, E.: 'Synergistic Selection': a darwinian frame for the evolution of complexity. J. Theor. Biol. **371**, 45–58 (2015)

Corning, P.A., Kline, S.J.: Thermodyanamics, information and life revisited, part II: Thermoeconomics and Control Information. Syst. Res. Behav. Sci. **15**, 453–482 (1998)

Davies, N.D., Krebs, J.R., West, S.: An Introduction to Behavioural Ecology, 4th edn. John Wiley, New York (2012)

Holland, J.H.: Emergence: From Chaos to Order. Addison-Wesley (Helix Books), Reading (1998)

Jablonka, E.: Epigenetic inheritance and plasticity: the responsive germline. Prog. Biophys. Mol. Biol. **111**, 99–107 (2013)

Jablonka, E., Raz, G.: Transgenerational epigenetic inheritance: prevalence, mechanisms, and implications for the study of heredity and evolution. Q. Rev. Biol. **84**(2), 131–176 (2009)

Jablonka, E., Lamb, M.J.: Evolution in Four Dimensions: Genetic, Epigenetic, Behavioral, and Symbolic Variation in the History of Life, Revised edn. MIT Press, Cambridge (2014)
Jablonka, E., Lamb, M.J., Avital, E.: 'Lamarckian' mechanisms in Darwinian evolution. Trends Ecol. Evol. **13**(5), 206–210 (1998)
Kauffmann, S.A.: At Home in the Universe: The Search for the Laws of Self-Organization and Complexity. Oxford University Press, New York (1995)
Laland, K.N., Sterelny, K., Odling-Smee, J., Hoppitt, W., Uller, T.: Cause and effect in biology revisited: is Mayr's proximate-ultimate dichotomy still useful? Science **334**, 1512–1516 (2011)
Laland, K.N., Odling-Smee, J., Hoppitt, W., Uller, T.: More on how and why: cause and effect in biology revisited. Biol. Philos. **28**(5), 719–745 (2013)
Lane, N.: Serial endosymbiosis or singular event at the origin of eukaryotes? J. Theor. Biol. **434**, 58–67 (2017). https://doi.org/10.1016/j.jtbi.2017.04.031
Love, A.C.: Rethinking the structure of evolutionary theory for an extended synthesis. In: Pigliucci, M., Müller, G.B. (eds.) Evolution – The Extended Synthesis, pp. 443–481. MIT Press, Cambridge, MA (2010)
McShea, D.W.: Bernd Rosslenbroich: on the origin of autonomy; a new look at the major transitions (book review). Biol. Philos. **30**(3), 439–446 (2015)
Michod, R.E.: Darwinian Dynamics, Evolutionary Transitions in Fitness and Individuality. Princeton University Press, Princeton, NJ (1999)
Michod, R.E.: Evolution of individuality during the transition from unicellular to multicellular life. Proc. Natl. Acad. Sci. **104**, 8613–8618 (2007)
Michod, R.E.: Evolutionary transitions in individuality: multicellularity and sex. In: Calcott, B., Sterelny, K. (eds.) The Major Transitions in Evolution Revisited, pp. 169–197. MIT Press, Cambridge (2011)
Morowitz, H.J.: Energy Flow in Biology. Academic Press, New York (1968)
Noble, D.: A Theory of Biological Relativity: No Privileged Level of Causation. Interface Focus **2**, 55–64 (2012)
Noble, D.: Physiology is Rocking the Foundations of Evolutionary Biology. Exp. Physiol. **98**(8), 1235–1243 (2013)
Noble, D., Jablonka, E., Joyner, M.J., Müller, G.B., Omholt, G.B.: Evolution evolves: physiology returns to centre stage. J. Physiol. **592**(11), 2237–2244 (2014). https://doi.org/10.1113/jphysiol.2014.273151/epdf
Massimo, P., Müller, G.B.: Evolution – The Extended Synthesis. MIT Press, Cambridge (2010)
Ratcliff, W.C., Ford Denison, R., Borrello, M., Travisano, M.: Experimental Evolution of Multicellularity. Proc. Natl. Acad. Sci. **109**, 1595–1600 (2012)
Ratcliff, W.C., Fankhauser, J.D., Rogers, D.W., Greig, D., Travisano, M.: Origins of multicellular evolvability in snowflake yeast. Nature Commun. **6** (2015). Article number: 6102. https://doi.org/10.1038/ncomms7102
Schneider, E.D., Sagan, D.: Into the Cool: Energy Flow, Thermodynamics, and Life. University of Chicago Press, Chicago (2005)
Schrödinger, E.: What is Life? The Physical Aspect of the Living Cell. Cambridge University Press, Cambridge, UK (1944)
Shapiro, J.A.: Revisiting the central dogma in the 21st Century. Ann. N. Y. Acad. Sci. **1178**, 6–28 (2009)
Shapiro, J.A.: Evolution: a View from the 21st Century. FT Press Science, Upper Saddle River (2011)
Torday, J.S.: Life Is simple—biologic complexity is an epiphenomenon. Biology **5**(2), 17 (2016). https://doi.org/10.3390/biology5020017
Torday, J.S., Rehan, V.K.: Evolution, The Logic of Biology. John Wiley, Hoboken (2017)

West-Eberhard, M.J.: Developmental Plasticity and Evolution. Oxford University Press, Oxford (2003)

Woese, C.R.: A new biology for a new century. Microbiol. Mol. Biol. Rev. **68**(2), 173–186 (2004)

Woese, C.R., Goldenfeld, N.: How the microbial world saved evolution from the scylla of molecular biology and the charybdis of the modern synthesis. Microbiol. Mol. Biol. Rev. **73** (1), 14–21 (2009)

Applications of Complex Systems in Socio-Economic Inequality Research: A Preliminary Survey

Czesław Mesjasz(✉)

Cracow University of Economic, Cracow, Poland
mesjaszc@uek.krakow.pl

Abstract. The ideas drawn from broadly defined systems thinking, including complex systems studies, have already been used to describe and explain socio-economic inequality, both directly and indirectly. Most of these works are good examples of applications of middle-range theories of socio-economic inequality, but many issues still need to be studied. The aim of this paper is to survey the different applications of complexity studies in the extant research on socio-economic inequality. Although inequality is a common phenomenon worldwide, attention in this paper is paid to the developed world, where poverty is not such a significant issue and where the 'information revolution' has had a greater impact.

Keywords: Complexity · Socio-economic inequality · Middle-range theories

1 Introduction

Socio-economic inequality studies are dominated by two approaches [35, 38, 41, 44]. In the first approach, empirical studies focus on describing the situation, often without a deeper explanation of the causes. In the second, inequality is seen within the framework of broad ideological and/or political considerations. A research gap exists, in which middle-range theoretical ideas may be placed. This gap can be at least partially filled with applications of complex systems studies.

The ideas drawn from broadly defined complex systems studies have already been employed directly and indirectly to describe and explain socio-economic inequality, e.g. [3, 14, 27]. The following questions could be proposed: What are the universal, systemic characteristics of inequality in social systems stemming from inherent differentiation (hierarchy, functional differentiation, clustering)? What are the consequences of the relative character of socio-economic inequality? What are the weaknesses of measuring socio-economic inequality? What is the specificity of socio-economic inequality in the modern 'Information Society'? Is there any '*eigendynamik*' of socio-economic inequality in a modern society (inequality as an 'emerging property')?

The aim of the present paper is to survey the extant research in socio-economic inequality and to show the different potential applications of complexity studies in this area. Although inequality is a worldwide phenomenon, more attention is paid here to

A. J. Morales et al. (Eds.): ICCS 2018, SPCOM, pp. 24–32, 2018.
https://doi.org/10.1007/978-3-319-96661-8_3

the developed world where poverty is not such a significant issue and where the 'information revolution' has a greater impact. A 'moderate constructivism' is proposed as the epistemological foundation of the paper. This foundation encompasses ideas from the philosophy of language, hermeneutics [32], linguistics [28] and constructivism [22, 40].

2 Interpretations and Definitions of Inequality in Society

Socio-economic inequality requires a precise interpretation since in multiple scholarly texts the notion of 'inequality' without any adjectives is used interchangeably with economic inequality, e.g. [33]. Inequality is considered in four key domains [11, p. 4]: socio-economic, health, political, and cultural.

Two distinctions need to be made in any analysis of inequality. The first is the difference between the unequal distribution of desirable life outcomes (health, happiness, educational success, or material possessions) and the unequal distribution of opportunities (access to power and life opportunities that facilitate the attainment of desirable outcomes). The second is the distinction between the unequal distribution of opportunities and outcomes between individuals and groups [11, p. 3]. The present paper focuses upon individuals.

Two interpretations of inequality in society can be distinguished – social and economic. Social inequality exists when resources and rights in a society are distributed unevenly, typically through norms of allocation that engender specific patterns along lines of socially defined types of individuals. They are differentiated according power, religion, kinship, prestige, race, ethnicity, gender, age, and class. Social rights include the labor market, source of income, health care, freedom of speech, education, political representation, and participation. Social inequality is held responsible for conflict, war, crisis, oppression, criminal activity, political unrest and instability, and indirectly it impacts upon economic growth.

Viewed as a sub-class of social inequality, economic inequality is depicted as 'the fundamental disparity that permits one individual certain material choices, while denying another individual those very same choices' [33, p. 1]. In a more specific way, economic inequality can be defined as the difference in various measures of economic well-being between individuals, in a group, between groups in a population, or between countries. It encompasses three areas: wealth, income and consumption. The issue of economic inequality is relevant to notions of equity, equality of outcome, and equality of opportunity [19]. It can be analyzed in a spatio-temporal framework, i.e. a territory, through the lens of various elements of social systems and across specific time scales.

3 The 'Hard' and 'Soft' Complexity of Social Systems

In his search for a definition of complexity, Lloyd [30] identified 45 interpretations of this term. An intricate picture of complexity studies emerges from the scheme proposed by Castellani [12]. Numerous definitions of complexity have been formulated (only a few are quoted here) [5, 7, 21, 24, 25, 34, 39, 46, 47, 48].

In both ‘first order cybernetics’ [2] and in ‘hard’ systems thinking [6] – where the observer is not taken into account – complexity was an important systemic feature. Systems complexity can be characterized by such traits as adaptability, adaptation, attractor, *autopoiesis*, bifurcations, the butterfly effect, chaos, a closed system, coevolution, complex adaptive systems, dynamical systems, edge of chaos, emerging properties, far-from-equilibrium states, a fitness landscape, fractals, hierarchy, non-linearity, the number of elements, the number of relations, an open system, path dependence, Power-Law, reflexivity, scale-free networks, self-organization, self-organized criticality, self-reflexivity, synergy, synergetics, turbulence, and viability.

Following [29], the above ideas can be defined as ‘hard’ complexity research by way of an analogy with ‘hard’ systems thinking, and embody ‘first order cybernetics’. This research includes mathematical modeling of systems with operationable and computable characteristics in physics, chemistry, natural sciences and in society. The ‘soft’ complexity research, also coined as an analogous term for ‘soft’ systems thinking [15] and ‘second order cybernetics’ [20], includes qualitative, often non-operationalizable and non-computable features of complexity. These ideas can be divided into two groups. The first group includes analogies and metaphors drawn from ‘hard’ complexity studies. They dominate in social sciences, though are often abused and misused. The second group includes qualitative indigenous concepts such as the complexity of the social systems of Luhmann [31] or ideas that are partly indigenous, and partly based on analogies and metaphors [16].

Subjectivity is the key feature of complexity in the ‘soft’ approach. In second-order cybernetics, or in a broader approach, i.e. constructivism [7, 22, 40], complexity is not an intrinsic property of an object but rather depends on the observer - ‘complexity, like beauty is in the eyes of the beholder’.

The complexity of social systems developed by Luhmann [31] is linked to reflexivity, self-reflexivity and self-reference, since a reduction in complexity is also a property of the system’s own self-observation, although no system can possess total self-insight. This phenomenon is representative of the epistemology of modern social sciences, where hermeneutics, learning, observation, self-observation, reflexivity and self-reflexivity, self-reference and subsequently intersubjectivity play an important role.

4 Internal Systemic Hierarchies and Socio-Economic Inequality

According to studies of society, hierarchical structures emerged as a consequence of the increasing complexity of decision making processes within primitive tribes since complex decisions could not be made by consensus. This could indicate a correlation between the increasing complexity of social systems and increasing inequality.

Hierarchies in social systems can emerge as a result of self-organization, although their emergence can also be stimulated. Hierarchy in terms of power and the allocation of resources can also be imposed [45]. Structural and functional differentiation in social systems encompasses the following: hierarchy, heterarchy and holarchy.

The hierarchy of social systems characterizes both ‘hard’ and ‘soft’ complexity. Although the internal and external hierarchies of systems have been studied, especially in the theory of control and automata, the universal ideas of Simon [42, 43] and Bertalanffy [6] continue to be applied in complex systems studies. In the simplest sense, hierarchy is a relation of subordination. Social systems include the following types of hierarchy:

(1) Power hierarchy.
(2) Functional hierarchy.
(3) Nested, recursive (fractal) hierarchy, which can be illustrated using the Russian "matrioshka" doll where the hierarchical elements are self-similar.
(4) Containment hierarchy is a nested hierarchy in which the subsets must differ from one another. Compositional containment hierarchy is an ordering of the parts that make up a system—the system is "composed" of these parts. Most engineered structures, whether natural or artificial, can be broken down in this manner. What is also important in this type of hierarchy the emerging properties occur, i.e. it is not possible to predict the behavior of a higher level system based upon observations of systems at a lower level of hierarchy.
(5) Subsumptive containment hierarchy - a classification of object classes.

Another type of internal ordering in systems is heterarchy, which involves a synthesis of any type of hierarchical ordering with elements of the same rank. Similarly to hierarchical structures, heterarchy may have a recursive character [23, pp. 127–157]. The next type of ordering affecting the internal differentiation of a society is holarchy. The concepts of holon and holarchy were introduced by Koestler [26]. A holon is an element of an entity, which is similar to the entity. Holons treated as elements of systems include information about the entire system. In complex adaptive systems (CAS) elements have only simple behavioral algorithms, e.g. bees in a beehive or simple models of human behavior [24, 46]. If elements are given cognitive capabilities, they could become similar to holarchic social systems. A holarchy can be also viewed as a kind of fractal structure.

The hierarchical structure of complex systems has been studied by Simon [42, 43]. He distinguished between subjective hierarchy deriving from cognitive limitations and physical hierarchy, and concluded that a hierarchical structure is an objective feature of any system [42]. Each level of a hierarchy has its specific characteristics. One of these characteristics that is important for socio-economic inequality is called near decomposability (nearly decomposable) [43].

According to Simon, all systems, - physical, social, biological, artificial – share the property of having a near decomposable architecture: they are organized into hierarchical layers of parts, parts of parts…., and so on, in such a way that interactions between elements belonging to the same parts are much more intense than interactions between elements belonging to different parts. By 'intense' interaction we mean that the behavior of one component depends more closely on the behavior of other components belonging to the same part than on components belonging to other parts. This property shows that elements at the same hierarchical level are naturally connected by stronger ties.

We may thus conclude and hypothetize that theories of the hierarchical, heterarchical and holarchical architecture of social systems are middle-range theories, applicable with a better understanding of the phenomenon of socio-economic inequality.

5 Complex Systems Models and Socio-Economic Inequality

Pareto had already made observations regarding wealth allocation within the universal framework of his 80–20 rule back in the early 20th Century [36, 37]. It was one of a number of attempts made to elaborate rigorous models of socio-economic inequality. These included not only methods of measurement and interpretation, e.g. the Gini coefficient, but also models allowing for the identification of causal links and predictions. Cumulative distributions with the Power-Law form are sometimes said to follow Zipf's Law or the Pareto distribution. Since Power-Law cumulative distributions imply a Power-Law form for p(x), 'Zipf's Law' and the 'Pareto distribution' are effectively synonymous with the 'Power-Law distribution'. Zipf's Law and the Pareto distribution differ graphically from one another [36, p. 4].

The Power-Law has become both a fundamental model of socio-economic inequality and a source of metaphors and analogies. The most significant feature of the Power-Law is that to some extent it reflects the situation in society, in which privileged groups receive the majority of resources. Of course, this picture is simplified but it reflects a certain rule existing in social systems. A distribution based on a Power-Law indicates that extreme events (or the richest people, or the biggest websites) account for most of the impact in that particular world, and everything falls off quickly afterwards.

The income distribution predicted by Pareto's law is skewed, with the frequency distribution declining monotonically, beginning at a minimum income. Pareto referred to this as the 'social pyramid'. This leads us to the issue of the connection between the hierarchical structure of most societies and their income distribution [37].

Empirical results of research in both 'hard' science and social studies led to the thesis that a shift can be observed from the Gaussian world to the Power-Law world [1]. This would mean that it is not the average which matters, but rather the rare and the unpredictable, which at the same time has a great impact on other elements of the system. However, as was proved by Clauset and co-workers, the Power-Law distribution is not so frequent as to legitimize generalizations of this kind [17].

Empirical studies found Power-Law behavior in the distribution of income in some countries [9, p. 2]. Using the empirical methodology for detecting Power-Law behavior introduced by Clauset [17], Brzeziński [9] found that top wealth values only follow Power-Law behavior in 35% of analyzed cases. Therefore, a complete empirical analysis would require conducting a statistical comparison of Power-Law models with other candidate distributions.

A number of other models are also applied in the study of socio-economic inequality: chemical kinetics-motivated Lotka-Volterra models, polymer physics-inspired models and, most importantly, models inspired by the kinetic theory of gases [14]. Applications of models, analogies and metaphors drawn from physics should be scrutinized and their importance in the study of socio-economic inequality should be re-assessed.

Scale-free networks elaborated by Barabási and Albert [3, 4] are also applied in studies of socio-economic inequality. Barabási and collaborators coined the term 'scale-free network' to describe the class of networks that exhibit Power-Law degree distribution. The 'preferential attachment' observed in scale-free networks may also be intuitively associated with inequality. This means that some objects of the network acquire more links since they acquire more links (the privileged are privileged, since they are privileged). However, the study [8] shows that scale-free networks are not too frequent either in nature or in social systems, and thus new studies are also required. The coincidence of scale-free network models with socio-economic inequality has been the subject of comments. Buchanan called random networks ‘egalitarian’ and scale-free networks ‘aristocratic’ [10, p. 119]. This coincidence was called the ‘rich get richer’ phenomenon [3, pp. 79–92; 10, pp. 106–120]. The hubs in scale-free networks were even called spiders in the net [19, pp. 150–151].

Even before the introduction of free-scale networks Castells [13] anticipated an increase in inequality in a network society. According to him, networks are characterized by a ‘space of flows’ that overwhelms and pervades the traditional ‘space of places’. Networks of capital, labor, information and markets are linked up through technology, valuable functions, people and localities around the world, and when switched off from their networks and territories these populations are deprived of values and interest in the dynamics of modern capitalism [13, p. 337; 19, p. 152].

Scale-free networks also have another important attribute enhancing their significance in any analysis of socio-economic inequality in the Information Society. In most cases where inequality is modelled with Power-Law models, the data reflect physical units. This means that ‘hard’ complexity is applied. In the case of networks more attention should be given to the constructivist character of complexity.

Network relationships in society are constructivist in form. There are two ways of constructing networks. First, an observer identifies relationships between behavior and/or the traits of various objects – individuals, groups – and puts them into a network. In this case, the network has at least a partly tangible foundation although its intersubjective character is also visible. The second type of network is purely intersubjective. For example, on the basis of my readings of a specific type of scholarly works, the search engine assigns me to a network of readers of literature on that topic. In this case the network is symbolic and, according to the previous considerations, we enter a world of intersubjectively created meanings. Thus, the challenges of ‘soft’ complexity come to the fore. In the Information Society, inequality in access to information, or in other words, to symbolic resources, could be even more differentiated. It could be hypothesized that the future Information Society will be affected by a deeper inequality, for example money is a form of information and a social construct as well.

6 Conclusions

The above survey of theories and empirical evidence allows us to draw a set of conclusions, which, due to the very idea of this paper, can also be viewed as areas of future research.

First and foremost, we must conclude that research into socio-economic inequality based upon various models drawn from complex studies has already provided some valuable insights into the characteristics of socio-economic inequality. This concerns in particular models based on the Power-Law and the scale-free networks associated with them.

A second more far reaching conclusion and at the same time hypothesis, or better, conjecture, can be formulated as follows: broadly defined complex systems studies are one of the most promising instruments for studying socio-economic inequality in a modern society. This applies to both ‘hard’ and ‘soft’ complexity. Complex systems ideas appear to be the most suitable in helping develop middle-range theories of socio-economic inequality.

Third, although the hierarchical structure of society is self-evident, the ideas drawn from complex systems studies provide additional evidence about the functioning of such structures. As a result, theoretical models and empirical evidence allow for a deeper understanding of the mechanisms giving rise to various types of inequality in social systems.

Fourth, the phenomenon of socio-economic inequality has different characteristics in an advanced Information Society compared with less developed countries. In the former, socio-economic inequality is affecting symbolic society (money is a symbol, the high value of equity of such companies as Facebook is to a large extent symbolic). In the latter, socio-economic inequality in most cases still concerns tangible aspects of social life. Studies of different types of socio-economic inequality demand different models be drawn from complex systems studies – inequality in reference to tangible and intangible assets.

References

1. Andriani, P., McKelvey, B.: From Gaussian to Paretian thinking: causes and implications of power-laws in organizations. Organ. Sci. **20**(6), 1053–1071 (2009)
2. Ashby, W.R.: An Introduction to Cybernetics. Wiley, New York (1963)
3. Barabási, A.-L.: Linked. How Everything is Connected to Everything else and What it Means for Business, Science, and Everyday Life. Penguin, New York (2003)
4. Barabási, A.-L., Albert, R.: Emergence of scaling in random networks. Science **286**(5439), 509–512 (1999)
5. Bar-Yam, Y.: Dynamics of complex systems. Addison-Wesley, Reading (1997)
6. von Bertalanffy, L.: General Systems Theory. Braziller, New York (1968)
7. Biggiero, L.: Sources of complexity. Hum. Syst. Nonlinear Dyn. Psychol. Life Sci. **5**(1), 3–19 (2001)
8. Broido A.D., Clauset, A.: Scale-free networks are rare. https://arxiv.org/abs/1801.03400. Accessed 16 Apr 2018
9. Brzeziński, M.: Do wealth distributions follow Power-Laws? Evidence from ‘rich lists’. https://arxiv.org/pdf/1304.0212.pdf. Accessed 14 Jan 2017
10. Buchanan, M.: Nexus: Small Worlds and the Groundbreaking Science of Networks. W.W. Norton & Company, New York (2002)
11. Carter, P.L., Reardon, S.F.: Inequality matters. A William T. Grant Foundation Inequality Paper, Stanford University, https://ed.stanford.edu/sites/default/files/inequalitymatters.pdf. Accessed 21 Jan 2017
12. Castellani, B.: Brian Castellani on the complexity sciences. http://theoryculturesociety.org/brian-castellani-on-the-complexity-sciences/. Accessed 20 Feb 2018
13. Castells, M.: The Information Age: Economy, Society and Culture: End of Millennium. Blackwell, Malden (1998)

14. Chatterjee, A., Ghosh, A., Inoue, J.-C., Chakrabarti, B.K.: Social inequality: from data to statistical physics modeling. J. Phys: Conf. Ser. **638**, 1–9 (2015)
15. Checkland, P.: Soft systems methodology: a thirty year retrospective. Syst. Res. Behav. Sci. **17**, 11–58 (2000)
16. Ciliers, P.: Complexity and Postmodernism. Routledge, London (1998)
17. Clauset, A., Shalizi, C.R., Newman, M.E.J.: Power-law distributions in empirical data. SIAM Rev. **51**(4), 661–703 (2009)
18. Economist: Measuring inequality. A three-headed hydra. http://www.economist.com/blogs/freeexchange/2014/07/measuring-inequality. Accessed 02 Nov 2017
19. von Foerster, H.: Observing systems. A Collection of papers by Heinz von Foerster. Intersystems Publications, Seaside, CA (1982)
20. Gell-Mann, M.: What is complexity? Complexity **1**(1), 16–19 (1995)
21. von Glasersfeld, E.: Radical Constructivism: A New Way of Knowing and Learning. The Farmer Press, London (1995)
22. Holland, J.D.: Hidden Order. How Adaptation Builds Complexity. Basic Books, New York (1995)
23. Kauffman, S.A.: At home in the Universe. The search for laws of self-organization and complexity. Oxford University Press, New York/Oxford (1995)
24. Koestler, A.: The Ghost in the Machine. Penguin Group, London (1967)
25. Krauss, A.: The scientific limits of understanding the (potential) relationship between complex social phenomena: the case of democracy and inequality. J. Econ. Methodol. (2015). https://doi.org/10.1080/1350178X.2015.1069372
26. Lakoff, G., Johnson, M.: Metaphors We Live By. University of Chicago Press, Chicago (1980/1995)
27. Lissack, M.R.: Complexity: The science, its vocabulary, and its relation to organizations. Emergence **1**(1), 110–126 (1999)
28. Lloyd, S.: Measures of complexity: a nonexhaustive List. IEEE Control Syst. Mag. **21**(4), 7–8 (2001)
29. Luhmann, N.: Social Systems. Stanford University Press, Palo Alto (1995)
30. Mantzavinos, C.: ‘Hermeneutics’. In: Zalta, E.N. (ed.) The Stanford Encyclopedia of Philosophy, Winter edn. https://plato.stanford.edu/archives/win2016/entries/hermeneutics/. Accessed 27 Dec 2016
31. McKay, A.: Defining and measuring inequality. Overseas Development Institute and University of Nottingham, Briefing Paper, 1, https://www.odi.org/sites/odi.org.uk/files/odi-assets/publications-opinion-files/3804.pdf. Accessed 13 Jan 2017
32. Mesjasz, C.: Complexity of social systems. Acta Phys. Polonica A **117**(4), 706–715. http://przyrbwn.icm.edu.pl/APP/PDF/117/a117z468.pdf. Accessed 17 Apr 2018
33. Milanovic, B.: Global Inequality. A New Approach for the Age of Globalization. The Belknap Press of Harvard University Press, Cambridge, MA (2016)
34. Newman, M.E.J.: Power-Laws, Pareto distributions and Zipf’s law. https://arxiv.org/pdf/cond-mat/0412004.pdf. Accessed 14 Mar 2015
35. Persky, J.: Retrospectives: Pareto’s Law. J. Econ. Perspect. **6**(2), 181–192 (1992)
36. Piketty, T.: Capital in the Twenty-First Century. Harvard University Press, Cambridge (2014)
37. Prigogine, I., Stengers, I.: Order out of Chaos. Bantam, New York (1984)
38. Searle, J.R.: The Construction of Social Reality. The Free Press, New York (1995)
39. Sen, A.K.: Inequality Re-Examined. Oxford University Press, Oxford (1995)
40. Simon, H.A.: The architecture of complexity. Proc. Am. Philos. Soc. **106**(6), 467–482 (1962)

41. Simon, H.A.: Near decomponsability and complexity: How a mind resides in a brain. In: Morowitz, J.H., Singer, J.L. (eds.) The mind, the brain and complex adaptive systems (Santa Fe Institute Series) 25-43. Addisson-Wesley, Reading, MA (1995)
42. Stiglitz, J.: The Great Divide. Unequal Societies and What We Can Do About Them. W. W. Norton & Company, New York (2015)
43. Turchin, P., Gavrilets, S.: Evolution of complex hierarchical societies. Social Evolution & History (8)2, http://www.socionauki.ru/journal/articles/129288/. Accessed 26 Jan 2017
44. Waldrop, M.M.: Complexity: The Emerging Science at the Edge of Order and Chaos. Simon & Schuster, New York (1992)
45. Weaver, W.: Science and complexity. Am. Sci. **36**(4), 536–544 (1948)
46. Wiener, N.: Cybernetics: Or Control and Communication in the Animal and the Machine. Hermann & Cie, Paris/MIT Press, Cambridge, MA (1948/1961)
47. Wittgenstein, L.: Philosophical Investigations. Blackwell Publishers, Oxford (2002)

Descartes, Gödel and Kuhn: Epiphenomenalism Defines a Limit on Reductive Logic

J. Rowan Scott(✉)

Department of Psychiatry, University of Alberta,
Suite 407, 11523-100 Avenue NW, Edmonton, AB T5K0J8, Canada
j.rowan.scott@gmail.com

Abstract. René Descartes' enduring contribution to philosophy, natural science and mathematics includes the unresolved residue of Cartesian dualism, as well as a singular 'bottom-up' interpretation of reductive logic sustained within the modern structure of the reductive natural science paradigm. Application of strong reductive logic leads to the perplexing *reductive epiphenomenalism* proposition.

Kurt Gödel's two famous incompleteness theorems provide an argument through analogy, demonstrating that *reductive epiphenomenalism of consciousness* is a logical and demonstrably *true* 'bottom-up' reductive proposition; characterized by conceptual paradox that cannot be resolved from inside the modern reductive science paradigm using sustained singular 'bottom-up' reductive logic. The argument by analogy concludes reductive epiphenomenalism is an *undecidable reductive proposition* declaring strong reductive logic to be *fundamentally incomplete*.

Thomas Kuhn's historical conception of a scientific revolution and modern explorations of contextual paradigm adaptation do not include descriptions of a limit on reductive logic associated with *reductive incompleteness*. One analogous implication of reductive incompleteness is the potential for an *unresolvable* and *undecidable reductive proposition*, stated in the paradigm and strong logic of reductive science, to become a *resolvable* and *decidable reductive proposition* within a closely related meta-reductive paradigm, preserving strong reductive logic but employing slightly different assumptions and premises. This opens the door to exploring functional adaptation of the reductive paradigm with the creation of adjacent possible meta-reductive paradigms.

Adjacent possible meta-reductive paradigms responding to reductive incompleteness, may be able to more closely mimic Nature's inherent evolutionary logic, provide novel solutions to unresolved or anomalous reductive scientific problems, and clarify the relationship formal reductive incompleteness might have with the natural logic of evolving systems.

Keywords: Descartes · Reductive logic · Reductive science · Reductionism
Reductive epiphenomenalism of consciousness
Unresolvable logical and conceptual paradox · Gödel
Incompleteness theorems · Undecidable reductive propositions
Reductive incompleteness · Kuhn · Reductive paradigm
Meta-reductive paradigm

A. J. Morales et al. (Eds.): ICCS 2018, SPCOM, pp. 33–52, 2018.
https://doi.org/10.1007/978-3-319-96661-8_4

1 René Descartes

René Descartes' enduring contribution to philosophy, natural science and mathematics includes the residue of Cartesian dualism associated with consciousness studies and the persistent application of a 'bottom-up' interpretation of reductive logic in modern science. The ongoing difficulty resolving the remaining brain/mind split may be directly related to Descartes' interpretation of reductive logic and the sustained application of his understanding within the modern structure of the reductive natural science paradigm.

René Descartes' initiated the consciousness debate with his famous pronouncement: *Cogito ergo sum* – "I think therefore I am." [21] Descartes' dictum declares sufficient, first-hand subjective knowledge of the existence of his own mind. Intangible and immaterial consciousness exists because he experiences it. Descartes' dualistic separation of fundamental substances, dividing *res extensa* (material substance that occupies space) from *res cogitans* (immaterial substance of the mind), ceded authority over immaterial spirit and mind to the Catholic Church and Inquisition, effectively excluding subjective consciousness and mind from the domain of scientific inquiry [80].

Descartes also suggested the material world must be approached with scientific skepticism, using objective observation and reductive analysis as the experimental and methodological tools for the nascent, historical natural sciences [3, 37]. Descartes' definition of *reduction* and his remarkably enduring initial formulation of reductive science appear first in his "*Discourse on the Method of Rightly Conducting the Reason, and Seeking Truth in the Sciences*". He composed four precepts, including, (1) *doubt*—scientific *doubt*; (2) *divide into parts*—*reduction*; (3) *ascend by little and little*—*synthesis*; and, (4) *enumeration*—declaring the essential importance of *mathematics* in science. Descartes envisioned a unified conception of a hierarchy of reductive natural sciences, bound together by the powerful and singular logic of reductive thought and the precise reasoning of mathematics [25]. A unified application of reductive logic metaphorically approaches phenomena with a 'downward' reductive focus, in order to decompose and understand the phenomena so that a subsequent 'bottom-up' synthesis can be achieved, often modeled with mathematical formalisms.

The underlying mechanisms or natural structures upon which any kind of unification of the sciences could be defined remains contentious [16]. However, in conformity with Descartes' solitary 'bottom-up' unifying interpretation of reductive logic, the modern natural sciences align themselves in a hierarchy from the most fundamental phenomena 'upward' toward more complex phenomena. Descartes' formulation of reductive science therefore survives in the effective modern reductive program, which proceeds in *three* steps: (1) *Reduction and Analysis*: Begin by taking apart a higher order phenomena into its disjoint elements and individually investigate these; (2) *Theoretical Formulation*: Using experimental evidence, imagination, and luck, formulate a model describing how the components relate and interact; (3) *Synthesis and Construction*: Using theory and experimental evidence, again compare the theoretical qualitative and quantitative success of the model with the experimental qualitative and quantitative behavior of the higher-level phenomena of interest, in order to demonstrate the scientific understanding of the phenomena is complete. Where possible then synthesize and construct the phenomena from its disjoint elements [70].

The application of reductive logic and the basic frame of the reductive science paradigm have not changed over four hundred years of vastly successful reductive exploration. With the exponential growth of scientific knowledge and increased understanding of cosmic evolution, biological evolution, self-organization and emergence in the hierarchy of complexity, it is worth reflecting on whether and under what circumstance, an adaptation of reductive logic and the structure of the reductive science paradigm, might better reflect the modern understanding of Nature.

2 Reduction

'*Reduction*' in modern science is associated with a reductionist claim that the logic of reductive thought and the epistemology of reductive science [79] mimic the ontology of Nature in specifiable ways [41]. The reductionist claim is justified by scientific evidence and the ongoing success of reductive science. The scientific reductive assertion states that the 'whole' can be *reduced* to the 'parts' constituting the 'whole', and the 'whole', including any emergent properties, can be fully accounted for by the 'parts', their 'interactions' and their 'relationships' [83]. The reductive understanding of '*fundamental'* is based on a belief that successful reduction of higher-order or more complex phenomena will ultimately arrive at an epistemological scientific description of an ontological substrate defining the most primitive components of the Universe [33]. In modern science, quantum physics is considered the most fundamental science describing the most fundamental and experimentally accessible entities, states and processes.

Nagel [57] composed an account of reduction as a kind of covering-law explanation, in which, one higher-level theory can be reduced to a second lower-level theory when it is possible to recognize that the theoretical terms of the first theory are related to or correspond to the theoretical terms of the second theory and it is possible to literally derive the first theory from the second. The reductive assumption that scientific disciplines and theories can correspond to one another in this way, allows reduction to serve as a framework for describing inter-level relationships and inter-level theories providing a route toward interdisciplinary integration. Nagel's account of covering theories suggests the inter-level relationship can be formally specified using reductive logic. The logic of the abstract relationship should not depend on either the content of the theories or the material structures the theories describe. Such clear and explicit logical and material correspondence between higher and lower theories is an abstract goal of reductive thought but is not that easy to achieve. This well-defined reductive goal has not, for instance, been accomplished in relation to the transition from quantum physics to relativity and Newtonian mechanics [50].

Reductive logic and the method of reductive science have been vastly successful, yet reductive epistemology does not quite capture Nature's ontological logic, as expressed in evolution or the self-organization of emergence in the hierarchy of complexity. Multiple lines of argument suggest reductive thought has limits and some authors offer partial remedial approaches to perceived shortcomings. In physics a broader view of emergence is suggested [51] or a more comprehensive narrative is recommended detailing the natural history of phenomena [82]. In evolutionary biology, beyond natural selection, an enhanced awareness of complex natural histories involving multiple intersecting causal factors is advocated [27]. In philosophy of mind [32] and the study of mind and consciousness [78], it is argued that there is something false, wrong or unfinished about the modern, reductive, materialist, neo-Darwinian scientific conception of Nature and consciousness. Further, it is difficult to articulate the boundary of the limitation in the absence of alternative conceptual frameworks [59].

Concern about the limits of reductive thought shape the mechanist perspective. The mechanist perspective offers an alternative to placing reduction at the conceptual center of natural science [22]. Mechanists criticize the conception that reduction should be assumed primarily to be a relationship between *theories*. For mechanists, scientists integrating their results are not simply building more elegant layered and corresponding theories; they are building theories about mechanisms. The mechanist perspective therefore tends to emphasize integrative pluralism in scientific research [54, 55]. Scientific achievements are collaborative and disjointed, adding incremental constraints to a developing representation describing how a mechanism works at one level and across levels [6, 22, 81]. Reductive logic and the mechanist perspective can work together and may ultimately belong in a novel paradigm placing entirely different concepts at the center of natural science.

Practical limits on reductive logic seem to be particularly relevant in relation to the often causally convoluted system dynamics in non-replicable or un-predictable settings associated with non-linear phenomena [69, 70, 84]. Some theorists go so far as postulating the existence of *non-reducible* hierarchically organized complex emergent phenomena [45, 46], in which case reductive logic is presumed to fail in part or entirely. Postulating non-reducible phenomena may be premature.

There may be unrecognized hard limits on the application of reductive logic in natural science. For instance, reductive logic is effective but may be limited *in principle* by a fundamental attribute of the logic that has not yet been recognized or successfully spelled-out. Understanding these unfamiliar limits and their implications needs to be addressed prior to advancing any concept or theory that rejects reductive logic in part or in whole.

3 Descartes, Reduction, Paradox and Epiphenomenalism

In *principle*, reduction works and reductive logic does not fail, despite *practical* difficulty conducting reduction in complicated contexts. Careful reductive analysis reveals even convoluted, non-linear, evolved, self-organized, hierarchically complex, emergent phenomena, such as consciousness, remain open to unyielding reduction [73]. However, among the anomalies associated with reductive logic, there are particular reductive conceptions suggesting the existence of hard limits.

Assuming Nature does not contain paradox, reductive theorists generally try to eliminate contradiction and paradox from scientific theory [85]. Despite the fact that reductive logic relentlessly works, particular well-composed reductive arguments using reductive logic nevertheless arrive at contradictory and paradoxical outcomes.

Careful dissection of reductive logic, seeking errors and omissions, as well as identifying partial or outright abandonment of reductive logic, often leads to clarification of paradox, contradiction or inconsistency; as well as providing deeper insights into the relationship between logic, science and natural domains [7]. Many scientific paradoxes and contradictions [8] can be resolved, when the source and structure of a conflict between observed phenomena, experimental results, theoretical constructions, mathematical models or narrative languages of description, can be determined.

Sometimes, however, no matter how hard scientists try, theorists are unable to extract or remove paradox from a particular reductive proposition. When a reductive proposition produces a paradox that cannot be resolved through careful examination of the logic and other identifiable sources of conflict, there may be something going on revealing a hard limit on reductive logic in natural science. For instance, in the hands of a careful and skeptical philosopher explicitly attempting to find a flaw in the strong reductive argument, William Seager's detailed logical, philosophical, scientific and mathematical analysis of the *reductive epiphenomenalism of consciousness proposition* and its *paradoxical* conclusion, reveals with some disappointment that the epiphenomenal argument is a very solid, step-by-step, logical and thus '*true*' reductive scientific proposal [73]. Seager carefully reviews all available historic and modern attempts to resolve the paradox of *reductive epiphenomenalism of consciousness* and finds no clear winner. The absurdly paradoxical conclusion of epiphenomenalism therefore must stand—*reductive epiphenomenalism* paradoxically insists that the complex phenomena and properties of subjective consciousness and causally impactful mind are only illusory epiphenomena of more fundamental quantum states [74]. It is this disturbing unresolved reductive paradox that may reveal the presence of a hard limit on reductive logic in science.

Thus, we see that René Descartes is responsible for two major modern philosophical and scientific conundrums. He initiated a four hundred year consciousness debate regarding his dualistic conception of material brain and immaterial mind, effectively pushing consciousness and mind into a bin of anomalies in relation to reductive natural science. Descartes is also responsible for the early formulation of unitary, 'bottom-up' reductive logic; a logic still applied in modern reductive science and the only available formal scientific logic used by natural science [23]. In the modern day, Descartes' conception of reductive logic arrives at a complex paradox associated with the reductive epiphenomenalism of consciousness proposition; which can be stated as a bizarre *self-referencing paradox* [8]—"I think *reductively*, therefore I am *not.*"

If the enigma of epiphenomenalism *can't be resolved*, the rigor of 'bottom-up' reductive logic *paradoxically erases consciousness* from the library of 'real' natural phenomena by translating consciousness into *illusory epiphenomena of quantum physics*. Descartes' complicated legacy challenges philosophers, scientists and mathematicians alike to find better solutions for both the residual Cartesian gap and the unresolved paradox of reductive logic and epiphenomenalism.

4 Kurt Gödel

'Reductive thinking', is a very special 'kind' of scientific and mathematical *'formal logic thinking'*. Kurt Gödel most succinctly defined a relationship shared by contradiction and paradox and formal logic, in his two famous incompleteness theorems published in 1931 [28–30]. A short summary of Gödel's important work provides an analogy and framework for confronting the contradiction and paradox at the center of *reductive epiphenomenalism of consciousness.*

There are four parts to Gödel's argument proving the two incompleteness theorems. First, there is Gödel's doctoral thesis, in which he learned about *formal completeness*. Then there is Gödel's exploration of *paradox*. This leads directly to the formulation and the precise, detailed, logical, proof of the *two incompleteness theorems*. Finally, there are the *implications of incompleteness* Gödel explores as a consequence of the incompleteness theorems. By responding to all four steps in the sequence of development of Gödel's thought, the analogy constructed links Gödel's work on incompleteness with reductive science and epiphenomenalism.

Gödel was interested in systems of abstract formal logic and mathematics. The components of an abstract formal system include, an 'alphabet', 'rules of grammar', 'axioms', 'rules of inference', definitions of 'grammatically well-formed statements', as well as derived and proven 'theorems'. The precise formal logic employed in a formal system is like a very finely tooled machine that leaves no space for fuzzy interpretation. The logic provides a mechanical procedure determining whether any given statement conforms to the system. To be useful, the logic of a formal system must be *consistent*. A formal system "is *consistent* if there is no statement such that the statement itself and its negation are both derivable in the system. Only consistent systems are of any interest in this context, for it is an elementary fact of logic that in an inconsistent formal system every statement is derivable, and consequently, such a system is trivially complete." [65].

Prior to beginning his work on *incompleteness*, Gödel began his academic career with a doctoral thesis, exploring *completeness* and the conditions in which an abstract formal system could be considered *closed, resolved, decidable* and *complete.* The mathematical ethos of Gödel's day was set by the agenda of David Hilbert [86] and a common belief that mathematics was on the verge of finalizing a formulation providing a complete understanding of modern mathematics [47]. Gödel began a journey that unraveled Hilbert's intention and the expectation of the mathematical community, when he focused on paradox and composed his detailed logical proof of two incompleteness theorems, which spell-out the conditions in which any formal system of sufficient complexity must be considered *open*, *unresolved*, *undecidable* and *incomplete* [64].

Gödel was interested in paradoxes of self-reference [11]. He began by drawing an analogy with two paradoxes of self-reference he specifically mentions; the 'Epimenides' or 'Liar's' paradox and antinomy [5], and 'Richards' paradox and antinomy [29, 38]. The 'Liar's' paradox creates a semantic analogue of Gödel's first incompleteness theorem, through a syntactical and abstract mathematical demonstration of an *undecidable proposition* within a formal system. 'Richard's' antinomy is a semantic antinomy that highlights the significance of logical *consistency* and the importance of differentiating clearly *mathematics* from *meta-mathematics*. This provides a semantic analogue for Gödel's second incompleteness theorem, through a syntactical, abstract mathematical demonstration of mathematical and meta-mathematical levels of argument.

The 'Liars Paradox' [13] is stated: "This statement is false." If "this statement is false" is true, then the sentence is false, but if the sentence states that it is false, and it is false, then it must be true, and one goes back and forth in the paradox. The paradox in part depends on the fact that English sentences can be constructed that cannot consistently or unambiguously be assigned a 'truth value', true or false, even though they are completely in accord with grammar and semantic rules.

The 'Liar's' Paradox is in part dependent upon an assumed binary decision that must be made between true and false—this is the assigned 'truth value'. From the vantage point of the paradox and the logic of binary decisions, the paradox cannot be resolved. From a meta-logical vantage point one can see that the ambiguity of English is a problem and binary decisions are not the only available kind of logical decision one might wish to consider in dealing with a statement created by a liar; nor are binary decisions the only way of thinking or dealing with a lie.

'Richard's' antinomy or paradox [58], starts with unambiguous real numbers and translates these precise mathematical objects into English statements. 'Richard's' paradox then results in an untenable contradiction between the level of unambiguous real numbers and the level of English statements about the numbers. The real number level and the English meta-level must be examined carefully to find the 'error'. The argument proceeds by demonstrating that a specific real number is unambiguous but the meta-level English statement describing the real number in natural language creates a statement defining the real number, for which there is no way to decide whether the meta-level English description is unambiguous or not. The paradox hinges on the realization that particular expressions of natural language describe real numbers unambiguously, while other expressions of natural language do not. While there is a way to demonstrate that real numbers are unambiguous, there is no way of determining unambiguously, which English statements unambiguously define a real number. The resolution of 'Richard's' paradox, then, is that there is no way to unambiguously determine exactly which English sentences are unambiguous when they define real numbers which also means there is no way to describe in a finite number of words whether an arbitrary English expression is a potential unambiguous definition of a real number [31].

The Halting problem is related to Richard's antinomy. Alan Turing, in studying computing, defined the Halting problem in 1936 in a response to Gödel's theorems. Turing proved that computational 'halting' is *undecidable* over all computing machines, meaning it is impossible to determine from the description of an arbitrary computer program and a given input, whether the program is going to finish running or run on for ever, thus one cannot determine from *inside* a computer program whether it is complete or forever incomplete [12]. Turing, aware of Gödel's incompleteness proof from 1931, was in search of significant examples that extended Gödel's very important result [4, 40, 47, 64].

Gödel drew an analogy from the Liar and Richard's paradox, since both are focused on *truth* and *falsehood*. Kurt Gödel constructed his analogy by shifting the attention of his readers toward a different paradox involving *truth* and *proof*. Gödel shifts attention from: "This sentence is *false*", to: "This statement is *not provable*" [14]. Gödel then proceeded to construct his 'logic bomb', which created a gulf between *truth* and *proof*, by shifting attention from the idea of *paradox* toward the mathematical concept of an *undecidable proposition. Undecidable propositions* are propositions that are "neither provable nor disprovable" [29, 30].

The *first* of Gödel's two theorems demonstrates the following: In abstract mathematical formal systems of sufficient complexity, which use a definable and consistent logic, there will inevitably be found *undecidable propositions* declaring the formal system to be *incomplete*. An *undecidable proposition* within a formal system is subtly different from a paradox or contradiction. Gödel's careful and detailed abstract logical argument creates a coding scheme that translates every statement, logical formula and proof, in Russell and Whiteheads' *Principia Mathematica* [43] into a mirror statement about natural numbers. He then takes the notion of 'truth' out of the Epimenides Paradox or the Liar Paradox, stated, "This statement is false" [14]; and replaces 'false' with an assertion about 'proof', in the form: "*This statement is not provable*". This statement about provability rather than true and false, is then also coded as an arithmetical, mirroring statement or counterpart, called a 'Gödel sentence' in abstract logic. He then coded the 'Gödel sentence', in a carefully detailed logical sequence, into the language of arithmetic [15]. The *informal* undecidable Gödel sentence is stated in the form, "This statement is not provable". The *formal* version of the Gödel sentence or statement, appears in Gödel's theorems in the form [*R*(*q*); *q*], creating a self-referential statement that asserts its own unprovability and declares its own undecidability [38]. "By focusing on provability rather than on truth, Gödel's sentence avoids the paradox. If formal arithmetic is consistent, meaning that only true statements can be proven, then Gödel's statement *must* be true. If it were false then it *could* be proven, contrary to the consistency! Furthermore, it cannot be proven, because that would demonstrate just the opposite of what it asserts, its unprovability!" [39].

Thus, in Gödel's first theorem, *truth* becomes separate from *proof*. It is possible for a statement to be grammatically correct, logical and consistent within the framework of a formal system and in this sense for the statement to be *true*. The same '*true'* statement may also be a statement that because of its own assertion about itself, *cannot be proven to be true or proven to be false*. This differentiation effectively separates *truth* from *proof*. A 'Gödel sentence' becomes a *true statement* and an *undecidable proposition* in a formal system, a statement that subtly stops short of paradox and contradiction by declaring itself, through self-reference, to be *undecidable*—"This statement *cannot be decided* as to its *truth* or *falsehood*". A Gödel sentence is logically *true* but it *cannot be proven to be true or false*. As long as the undecidable Gödel statement is not forced into paradox and contradiction through decision, it remains part of a logical system and in effect *protects* the logical system from the dangers of paradox, contradiction and inconsistency. Gödel's first theorem goes on to prove that every version of the 'Gödel sentence' in every conceivable formalization of arithmetic must be '*true*' if the formal systems are sufficiently complex and consistent [64].

Consistency is carefully addressed in Gödel's second incompleteness theorem. As long as the Gödel statement located in a formal system is left *undecided* the statement can be logical and true but cannot be proven to be true or false within the formal system and its rules. If the statement is *decided* to be *true* or *false,* the statement then becomes a *contradiction* or *paradox,* threatening the logic and the consistency of the formal system. The idea of an *undecidable proposition* or 'Gödel sentence' forms a bridge between Gödel's *first* and *second* incompleteness theorems. The *first* theorem is focused on *undecidable propositions* in relation to the 'logic', the 'truth' and the 'provability' of statements within a formal system. The *second* incompleteness theorem is focused on *undecidable propositions* in relation to '*proof*' of '*consistency*' of statements within a formal system.

Gödel's *second* incompleteness theorem demonstrates that the logic of a formal system, when it is sufficiently strong enough that it contains *undecidable propositions*; is also too weak, or not strong enough, to 'prove' its own 'consistency'. Therefore, the *consistency* of a formal system cannot be determined from inside the system itself—*consistency* must be approached through *meta-consideration* derived outside the particular formal system and its logic [15]. "Gödel showed that if the consistency of the formal system could be demonstrated inside the system itself, then the informal argument just given could be formalized and the formalized version of the statement, "This statement is unprovable," would itself be proven, thereby contradicting itself and demonstrating the inconsistency of the system!" [39]. Consistency, therefore, *must* be determined from outside a formal system, in order to once again protect the system from illogic, contradiction, paradox and inconsistency. The necessity of meta-consideration, in relation to determining the consistency of a formal system of sufficient complexity, reveals another way in which abstract pure mathematical formal systems of sufficient complexity, *cannot be closed, resolved, decided or complete* structures.

Gödel goes on to prove two further significant implications of the incompleteness theorems. *First,* what must remain an *undecidable proposition* or 'Gödel sentence' in *one* abstract formal mathematical system, can often *be decided* and have a significantly different meaning and implication within a closely related formal system using a slightly different set of axioms and theorems [13, 56, 58, 64]. *Second*, Gödel's work additionally reveals that in an alternative formal system composed expressly to get around the presence of an *undecidable proposition* in *one* formal system, by adding axioms or theorems and by successfully making the 'Gödel sentence' provable or decidable: In such an alternative formal system, the system will inevitably run into its own *unprovable* 'Gödel sentence' or *undecidable* proposition [15, 47, 64].

5 Gödel and Reductive Epiphenomenalism of Consciousness

It is conceivable that reductive logic and reductive science create a sufficiently complex system that it is possible to construct or discover a logically *'true'* reductive theoretical proposition stating a linguistically '*unresolvable'* and formally '*undecidable'* self-referencing contradiction or paradox.

Is *reductive epiphenomenalism of consciousness* such an *unresolvable* reductive scientific proposition? More specifically, by analogy with Gödel's four-part proof of his two incompleteness theorems, is *reductive epiphenomenalism of consciousness* a *logical, true*, but also *unresolvable* and *undecidable reductive* proposition? If *reductive epiphenomenalism of consciousness* is a *true* but *undecidable* reductive proposition, as in the first of Gödel's theorems, it must be *left undecided* within the reductive formal system of logic, in order to *protect* the reductive logical system from *illogic, contradiction, paradox*, and *inconsistency*. Further, as in the second of Gödel's theorems, in order to avoid contradiction and paradox it may also be necessary that the *consistency* of reductive logic, just as in abstract formal systems, must be determined through meta-consideration, demonstrating again that the *reductive system of logic cannot be contained, closed, resolved, fully decided* or *complete*. By analogy, if r*eductive epiphenomenalism of consciousness* is an *unresolvable* and *undecidable* reductive statement, it effectively declares reductive logic and the reductive science paradigm to be a system employing a formal system of logic of sufficient complexity that it exhibits *formal logical reductive incompleteness*. It is worth a closer look at the epiphenomenalism proposition.

Reductive epiphenomenalism of consciousness is a reductive proposition employing strong reductive logic that reduces mind to brain, brain to material body, and material body to the most fundamental physical description possible in natural science; a description composed in the language of quantum physics. The *reductive epiphenomenalism of consciousness proposition* implies the conclusion that mind and all its manifestations, including subjective awareness and experience, including the qualia of perception [17–19], including the experience of freewill and willful and intentional action [36], including meaningful subjective emotional experience, including any sense of causal authority as an active agent composing aspects of our own environment in interactions and relationships in the world [35]; *all* these properties and all other defined properties of mind and consciousness are entirely transformed into an *illusion* and a *fantasy,* mere *epiphenomena* of a fundamental physical quantum description [73]. As a generalization, the *epiphenomenalism proposition* [66] reduces *all* evolved, emergent, hierarchically organized complexity, all higher-level causal interaction and relationships, whether associated with consciousness or not, *fully* and *totally* to a quantum description wherein all causal explanation ultimately resides in the quantum account.

The '*true'* logic of reductive epiphenomenalism finds experimental '*proof*' and support in the work of Libet [52]. Taking the true logic and experimental proof to heart, Harris [36] and Harari [35] try to comprehend the absurd meaning for human existence of the paradoxical illusion created by epiphenomenalism. Meanwhile, other corners of the philosophical community often sidestep epiphenomenalism through a variety of adaptations of reductive logic or by altering the relationship consciousness has with the rest of reality [74]. Concurrently, most of the scientific community try and ignore the theoretical conflict, contradiction and paradox posed by epiphenomenalism while they continue the study of consciousness 'as if' it is a valid phenomena rather than an illusion.

The vast enterprise of functional neuroscience, including interpersonal neuroscience [76, 77], often *accept 'as a given'* causally significant brain, mind and consciousness. *Assuming* causally effective consciousness, research then paradoxically proceeds with a reductive scientific approach to function, effectively trying to reduce mind and complex conscious phenomena to manifestations of brain [9]. In the case of interpersonal neuroscience, mind is something interpersonal, more than, not equivalent to, and not reducible to brain [78]. This entire enterprise, in one way or another, abandons, partially implements or alters the reductive logic of natural science. Functional neuroscience and interpersonal neuroscience sidestep the strong reductive logic of the epiphenomenal argument and its further implications, mostly on the basis that its implications are absurd and not because of any careful validation of logical failure or demonstrated experimental falsification.

Popper's criteria for a *scientific epistemology* [62] state that in order to be called "scientific", hypotheses have to be *verifiable* and *falsifiable*. Thus, hypotheses must be both *logically* and *experimentally* accessible and not merely descriptive in nature [61]. In light of Popper's criteria, there are a number of inter-related postulates, premises, assumptions and logical, experimental, mathematical and narrative considerations, that must be accounted for, in order to assess the *logical* and *scientific* merit of the epiphenomenalism proposition and in order to determine whether or not it is *possible* to unravel its' contradictions and paradoxes.

Start from the *postulate*; reductive science must avoid contradiction and paradox in scientific theory [85]. Accept the *premise*; ontologically neutral *reductive logic works* in natural science [16, 83]. Permit the *possibility*; *beyond* the known *practical limits* of reductive logic, the epistemology [79] of applied *reductive logic may have unrecognized, in principle, limits* [41]. Admit the *assumption*; *reductive epiphenomenalism of consciousness*, as a *hypothesis*, is a *logically true* reductive scientific proposition with some experimental support that has never been found to contain any kind of logical flaw [73]. Then, consider the following:

1. The outcome of *proving* reductive epiphenomenalism *true* is a *complex reductive paradox of self-reference* [8]—consciousness *is and is not*. The paradox can be stated in two parts: "A *conscious participatory* and *intentional* human *mind* composes the *logically true statement of reductive epiphenomenalism of consciousness*: When the *true* statement is *proven true*, the conscious mind and author of the statement are, by direct implication, obliterated from reality, as the statement *paradoxically* translates consciousness and mind into illusory quantum *epiphenomena*".
2. In a process of verification or falsification through observation and experimental evidence [34], *presume* a *necessary* weight of observational and experimental evidence accumulates that it is considered *sufficient* [10] to declare the epiphenomenal proposition *experimentally proven to be true*. The entire community of mindful, conscious, willful, participatory scientists conducting the scientific experimental exercise are then an *illusion, paradoxically erased* from the Universe and transformed into *quantum epiphenomena* [67].
3. Some philosophers and theorists suggest reductive logic should only be used judiciously, in particular limited contexts, in order to avoid contradiction, paradox and the 'excessive claims' of strong reductive logic, such as postulated by reductive epiphenomenalism [23]. However, this diluted approach sidesteps important points:
    a. Partially watering down or totally abandoning reductive logic is a bad idea—it's the only formal logic natural science has at its disposal. Formal reductive logic,

its paradoxes *and its limits* should therefore be spelled-out before any adaptation of the logic should be considered.

b. Science intends to model Nature and Nature should be consulted in order to clarify how far reductive logic can be trusted in the task of approximating Nature. This exploration might prove strong reductive logic too weak to model Nature.

4. Reductive science might accumulate a large enough body of observation, theoretical and experimentally verifiable *contradictory* evidence that it effectively *proves* the epiphenomenal proposition *theoretically* and *experimentally false*—for example: fundamental physics may *enable* the existence of consciousness and mind, but quantum physics *may not limit* the causal power of hierarchically organized complex consciousness and emergent mind [26]. In this case, Nature gets the last word, because science is after all about attempting to understand and approximate Nature. If this does happen, reductive logic would remain logical and internally consistent, but it would be in deep trouble in relation to its scientific task. Faced with this 'minor catastrophe', formal reductive logic, at best, would become a weak mimic or a good approximation of a very limited aspect of natural evolutionary 'logic'—at worst, it might need to be abandoned entirely, to be replaced by what?

   a. Hypothetically, exploration of the relationship between, reductive logic (its defined and as yet unrecognized limits) *and* Nature (with its defined and as yet unrecognized evolutionary pattern), might discover, with surprise, a natural evolutionary pattern sharing similarities with the structure and application of reductive logic *and* its limits. A symmetric pattern might only become visible if strong reductive logic is sustained and reductive incompleteness is understood. Reductive logic might then *conform* to a repeating configuration in Nature and natural evolution!

5. Imagine a circumstance in which an annoyed, conscious, willful, reductive scientist, smugly succeeds in creating *a rigorously formal, logical, true and proven to be true, reductive argument* demonstrating that consciousness and mind exist as *more* than mere quantum epiphenomena—blatantly *contradicting* the formal logic and implication of the reductive epiphenomenalism of consciousness proposition. Such a contradictory argument, should it ever exist, would be a 'serious logical catastrophe' for reductive science. If *both contradictory* arguments are *true* and *proven true,* then reductive logic, through meta-level consideration, is an *inconsistent form of logic*—capable of arriving at *true* and *false* statements in relation to the *same* proposition and the same phenomena. Reductive logic in science is then a *trivial* logic capable of proving anything and incapable of differentiating true from false.

Thus, in all five circumstances focused on reductive logic and the scientific method, reductive logic is in trouble! The least troubling option above is the 'minor catastrophe' in which reductive logic must be supplemented with some alternative approach in order to more closely mimic Nature and evolution. William Seager demonstrated there are presently no clear philosophical winners capable of unraveling the paradox of epiphenomenalism [73]. The present discussion demonstrates there are presently no clear formal logic or scientific resolutions available either—*reductive epiphenomenalism of consciousness* creates contradiction and paradox that *cannot be resolved* using reductive logic. Therefore, the paradox of reductive epiphenomenalism *cannot be resolved* from *inside* the paradigm of modern reductive natural science.

Overall, to *avoid paradox, contradiction* or *logical inconsistency*, the *unresolvable* reductive epiphenomenalism of consciousness proposition must be left *undecided*. It must stand as an *undecidable reductive proposition* declaring reductive logic and reductive science to be *a sufficiently complex system of logic and thought capable of formal incompleteness*. *Truth* and *proof* in reductive natural science can be separated in reductive logic just as in formal logic and abstract formal systems. The undecidable epiphenomenalism proposition defines a *hard limit* on the application of reductive logic in the modern reductive science paradigm.

6 Thomas Kuhn: Epiphenomenalism Resolved and Decided

Thomas Kuhn [48, 49] envisioned periods of normal science in which scientists focus on problems considered acceptable and accessible, within the structure of a given theoretical frame or paradigm. The frame of normal reductive science naturally leaves some problems in categories defined as unsolved, unacceptable, or anomalous. Kuhn's initial statement defining the structure of scientific revolutions used 'theory' and 'paradigm' as interchangeable terms, leaving somewhat vague the potential scale of scientific transformations and revolutions. Modern work has extended Kuhn's exploration of scientific transformative events, studying lesser and greater scales of adaptation or revolution. When science enters into a period of upheaval on lesser scales, theories and methods can be modified and adapted to new contexts, but in particularly profound cases, on a much larger scale, an entire theory or scientific paradigm can be put aside in the face of novel ideas and successful alternative theoretical or paradigm structures [60].

The general framework of the reductive science paradigm and the 'bottom-up' application of reductive logic have remained in place for four centuries of successful scientific exploration. Descartes' historical 'bottom-up' conception of reductive science and the singular 'bottom-up' application of reductive logic in modern science has never faced a conceptual revolution of the kind or scale envisioned by Kuhn. However, a large library of 'unacceptable', unsolved or anomalous problems has accumulated within the reductive paradigm and a major paradigm shift or transformation of reductive science may be on the horizon.

The likely future structure of a potential paradigm shift and transformation of reductive science can be spelled-out through a further analogy with Kurt Gödel's work. Most important among Kurt Gödel's stated implications, is the general expectation that *an undecidable proposition in one formal system might be found to be resolvable and decidable, in a closely related formal system using slightly different axioms and assumptions* [64].

The history of enigmas and anomalies in mathematics and reductive scientific theory contains examples of statements and propositions posing conflict, contradiction and paradox that on careful examination are found to be *unresolvable* and *undecidable* in an existing frame of reference. By analogy, within a different mathematical or scientific frame of reference using slightly different premises, the unresolvable and undecidable propositions may prove to be *resolvable* and *decidable*. The ultimate resolution and decision of these unresolvable propositions in the past has depended on imagining or discovering an altered frame of reference creating a new meaning for the unresolved proposition, often transforming and opening whole domains of mathematics and science.

Two examples come to mind: *First*, Pythagoras and his conception of a 'closed universe' of whole numbers, which led to a confrontation with the unresolvable, undecidable and incomprehensible number $\sqrt{2}$. The resolution took the form of recognizing the incompleteness of whole numbers and the discovery of irrational numbers. This event opened up the vista of modern mathematics. *Second*, Isaac Newton's incredibly effective gravitational equation is very successful as an approximation but an incomprehensible mechanism it proposed, requiring *instantaneous action at a distance* formulated in the assumed deterministic fixed space and invariant time of Newtonian mechanics. This was considered to be a disturbing, incongruous, unresolvable and undecidable proposition. Newton's "great absurdity" and unsolvable puzzle was ultimately resolved and decided by Einstein in the form of special and general relativity theory [85].

The Pythagorean and Newtonian examples reveal an abstract pattern of mathematical and scientific transformation, not fully envisioned by Thomas Kuhn or any recent theorist, but implicit in Kurt Gödel's stated implications of his incompleteness theorems. The abstract pattern involves, recognition of potential paradox, subtle avoidance of paradox, the introduction of an undecidable proposition, generalization across multiple possible or similar formal system structures, as well as potential resolution of the undecidable statement within an altered system, frame of reference or paradigm structure. *Reductive epiphenomenalism of consciousness* exemplifies a troubling and unsolved modern paradox that can be subtly formulated as an *unresolvable* and *undecidable* proposition within the reductive paradigm. If, the reductive paradigm, reductive logic and its successes, can then be *preserved* and *encompassed as a special case* within an adjacent possible meta-reductive paradigm (MRP) using slightly different premises and assumptions, then, the unresolvable paradox and undecidable proposition of epiphenomenalism might become resolvable and decidable within the altered paradigm, with very different meaning and implication.

Such a successful adapted meta-reductive paradigm structure might entail *many* inter-locked, closely related but slightly different premises and assumptions, suggesting novel interpretations of natural phenomena and the complexity [71] explored by meta-reductive science [72]. In the present context, we are searching, at a minimum, for *one* adapted premise or assumption that offers an approach in which a novel adjacent possible meta-reductive paradigm can respond to the specific, unresolvable and undecidable, reductive epiphenomenalism proposition, by finding a path to its resolution, decision and novel meaning within a meta-reductive frame.

Imagine one slightly altered premise and assumption that could resolve the paradox and substantially alter the meaning of reductive epiphenomenalism. An imagined premise involves abandoning Descartes' and modern sciences' historical premise of singular 'bottom-up' reductive logic and replacing it, at a minimum, with an adapted premise; recognizing a need for *dual reductive description* involving 'bottom-up' causation and 'top-down' causal influences [26]. Implement the adapted premise in the context of evolved, hierarchically emergent [42] complex systems, or in the context of a complex adaptive system (CAS) model describing consciousness and the boundary of brain and emergent mind in a social context [53]. Such an adapted 'dual description' premise has been suggested and applied in the context of a CAS model of consciousness, in a psychiatric and psychotherapy context involving Dynamical Systems Therapy [75].

An adapted premise encompassing the reductive paradigm in a synthesis of dual reductive 'bottom-up' and 'top-down' descriptions in a meta-reductive paradigm (MRP), resolves the paradox of epiphenomenalism and decides in favor of a different meaning for reductive epiphenomenalism of consciousness. In a dual meta-reductive paradigm, the epiphenomenal proposition no longer stands as a singular, complete, rigorously logical 'bottom-up' reductive proposition. The 'bottom-up' brain can influence the mind and the 'top-down' mind can influence the brain [68, 76–78].

Deciding the meaning of the reductive epiphenomenal proposition takes the form of deciding in favor of a novel interpretation of the proposition based on the epistemology of a dual meta-reductive paradigm. Reductive logic must be applied twice and the interpretive meaning moves away from paradoxical epiphenomenal absurdity toward a narrative in which fundamental physics enables causally efficacious, hierarchically organized complexity [26]. In an encompassing complex dual MRP scientific narrative, reductive epiphenomenalism becomes an incomplete and partial statement composed by a historic and incomplete reductive natural science. Subjective conscious experience and the sense of personal agency, personal will power and intentionality can be reinterpreted in light of the dual causal description created by a dual MRP epistemology.

The adapted premise of a dual MRP responds to reductive incompleteness and responds to a reductionist blind spot [2]. In addition, the dual premise ultimately provides a more complete and resolved statement and narrative in which consciousness and emergent mind possess real emergent causal power and can do real emergent work [1]. Dual reductive accounting in a MRP significantly alters the basic understanding and theory of emergence in natural evolution and complexity [20]. While a dual MRP may not yet address all the philosophical obstacles to a modern science of consciousness [24], the multiple descriptions required in a MRP ultimately provide a better approximation of Nature's evolved, self-organized and emergent complexity.

An adapted premise demanding *dual* 'bottom-up' and 'top-down' multiple reductive descriptions might not be restricted to hierarchically organized complex models of CAS. The adapted premise might more generally make a demand for *dual* 'bottom-up' and 'top-down' multiple reductive descriptions throughout the entire hierarchy of sciences encompassed by the meta-reductive paradigm [26]. This adaptation of a basic premise in a meta-science might then have implication for resolving a number of other unresolved scientific problems. For instance, the lack of correspondence between quantum physics and relativity theory [50], and the incommensurate realities composed by quantum physics and relativity theory [63], might require a fundamental shift, from 'bottom-up' reductive interpretations, toward *dual* 'reductive thinking' and dual 'bottom-up' and 'top-down' reductive synthesis. It may be necessary to develop a generalized understanding of *dual* 'reductive thinking' integrated with 'emergent thinking', throughout all fundamental sciences and the entire hierarchy of sciences, up to and including the neurosciences studying brain and emergent mind [51].

The consequence of further adaptations of other premises and assumptions in a more complex meta-reductive paradigm is worth considering.

7 Conclusion

The discovery of one undecidable reductive proposition defines reductive logic and the reductive paradigm as sufficiently complex and susceptible to formal reductive incompleteness. The implications of reductive incompleteness associated with reductive epiphenomenalism of consciousness, opens a window on a panorama of novel adjacent possible meta-reductive paradigms, involving one or many altered and adapted premises and assumptions.

The existence of reductive incompleteness raises many further unanswered questions. What makes a theory or a paradigm sufficiently complex that it contains necessary incompleteness? Can incompleteness be generalized across all sciences? Does Nature instantiate natural forms of 'incompleteness' in the sequence of change, evolution, and the self-organization of the hierarchy of complexity?

It is possible for reductive science to compose, closed, resolved, decided and complete theoretical frameworks. By analogy, consistent with Gödel's work on completeness and his work on the two incompleteness theorems, any consistent reductive scientific theory of sufficient complexity will reveal the presence of undecidable reductive propositions and reductive incompleteness. When sufficiently complex, reductive scientific theory, the reductive paradigm and ultimately any meta-reductive paradigm created precisely to respond to identified incompleteness, will inevitably reveal, they too are open, contain unresolvable paradox, formulate undecidable propositions, and are fundamentally incomplete.

Acknowledgments. Dr. Norman Davies, Dr. Sol Levin, Dr. William Dewhurst, Dr. James McMullin, Dr. Jack Tuszynski, Dr. Don Page, and the steadfast support of Dr. Yakov Shapiro. Any errors or omissions in this paper are entirely my own responsibility.

References

1. Abbott, R.: Emergence explained: abstractions: getting epiphenomena to do real work. Complexity **12**(1), 13–26 (2006)
2. Abbott, R.: The reductionist blind spot. Complexity **14**(5), 10–22 (2009)
3. Andersen, H., Hepburn, B.: Scientific method. In: Zalta, E.N. (ed.) The Stanford Encyclopedia of Philosophy, Summer edn. Stanford Metaphysics Research Laboratory, Stanford (2016). https://plato.stanford.edu/archives/sum2016/entries/scientific-method/
4. Barker-Plummer, D.: Turing machines. In: Zalta, E.N. (ed.) The Stanford Encyclopedia of Philosophy, Winter edn. Stanford Metaphysics Laboratory: Stanford University (2016). https://plato.stanford.edu/archives/win2016/entries/turing-machine/
5. Beall, J., Glanzberg, M., Ripley, D.: Liar paradox. In: Zalta, E.N. (ed.) The Stanford Encyclopedia of Philosophy, Fall edn. Stanford Metaphysics Laboratory: Stanford University (2017). https://plato.stanford.edu/archives/fall2017/entries/liar-paradox/

6. Bechtel, W.: Looking down, around, and up: mechanistic explanation in psychology. Philos. Psychol. **22**, 543–564 (2009)
7. Bird, A.: Thomas kuhn. In: Zalta, E.N. (ed.) The Stanford Encyclopedia of Philosophy. Stanford Metaphysics Laboratory: Stanford University (2013). https://plato.stanford.edu/archives/fall2013/entries/thomas-kuhn/
8. Bolander, T.: Self-reference. In: Zalta, E.N. (ed.) The Stanford Encyclopedia of Philosophy, Fall edn. Stanford Metaphysics Laboratory: Stanford University (2017). https://plato.stanford.edu/archives/fall2017/entries/self-reference/
9. Boleyn-Fitzgerald, M.: Pictures of the Mind: What the New Neuroscience Tells Us About Who We Are. Pearson Education Inc. published as FT Press, Upper Saddle River, New Jersey (2010)
10. Brennan, A.: Necessary and sufficient conditions. In: Edward, N.Z (ed.) The Stanford Encyclopedia of Philosophy, Summer edn. Stanford Metaphysics Laboratory (2017). https://plato.stanford.edu/archives/sum2017/entries/necssary-sufficient/
11. Cantini, A.: Paradoxes and contemporary logic. In: Zalta, E.N. (ed.) The Stanford Encyclopedia of Philosophy, Fall edn. Metaphysics Lab, Stanford Univeristy, Stanford (2014). https://plato.stanford.edu/archives/fall2014/entries/paradoxes-contemporary-logic/
12. Casti, J.: The Halting Theorem (Theory of Computation). In: Five Golden Rules: Great Theories of 20th Century Mathematics–and Why They Matter, pp. 135–180. John Wiley and Sons Inc., New York (1996)
13. Casti, J., DePauli, W.: Gödel: A Life of Logic, pp. 48–52. Perseus Publishing, Cambridge (2000)
14. Casti, J.L., DePauli, W.: Godel: A Life of Logic, pp. 48–49. Perseus Publishing (2000a)
15. Casti, J.L., DePauli, W.: Gödel: A Life of Logic. Perseus Publishing (2000b)
16. Cat, J.: The unity of science. In: Zalta, E.N. (ed.) The Stanford Encyclopaedia of Philosophy, Fall edn. (2017). https://plato.stanford.edu/archives/fall2017/entries/scientific-unity/
17. Chalmers, D.: Facing up to the hard problem of consciousness. J. Conscious. Stud. **2**, 200–219 (1995)
18. Chalmers, D.: The Conscious Mind. In Search of a Fundamental Theory. Oxford University Press, Oxford (1996)
19. Chalmers, D.: Facing up to the problem of consciousness. In: Shear, J. (ed.) Explaining Consciousness, pp. 9–32. MIT Press, Cambridge (1997)
20. Corning, P.A.: The re-emergence of emergence: a venerable concept in search of a theory. Complexity **7**(6), 18–30 (2002)
21. Cottingham, J.G.: Descartes. In: Gregory, R.L. (ed.) Oxford Companion to the Mind, p. 189. Oxford University Press, Oxford (1988)
22. Craver, C., Tabery, J.: Mechanisms in science. In: Zalta, E.N. (ed.) The Stanford Encyclopedia of Philosophy, Spring edn. Stanford Metaphysics Laboratory: Stanford University (2017). https://plato.stanford.edu/archives/spr2017/entries/science-mechanisms/
23. Dennett, D.C.: Darwin's Dangerous Idea: Evolution and the Meanings of Life, pp. 80–85. Simon and Schuster, New York (1995)
24. Dennett, D.C.: Sweet Dreams: Philosophical Obstacles to a Science of Consciouness. A Bradford Book, The MIT Press, Cambridg, Massachussetts (2005)
25. Descartes, R.: Discourse on the Method of Rightly Conducting the Reason, and Seeking Truth in the Sciences: Gutenberg EBook Project (1637)
26. Ellis, G.: How Can Physics Underlie the Mind: Top-Down Causation in the Human Context. Springer, Heidelberg (2016)
27. Fodor, J., Piattelli-Palmarini, M.: What Darwin got wrong, pp. 153–163. Farrar, Straus and Giroux Publishers, New York (2010)

28. Gödel, K.: Über formal unentscheidbare Sätze der Principia Mathematica und verwandter Systeme. Monatshefte für Mathematik und Physik **38**, 173–198 (1931)
29. Gödel, K.: On formally undecidable propositions of principia mathematica and related systems (B. Meltzer, Trans.). In: Braithwaite, R. (ed.) Electronic Reprint Edition, Introduction by D.R. Hofstadher. Basic Books, New York (1992)
30. Gödel, K.: On formally undecidable propositions of principia mathematica and related systems. Braithewaite, R.B. (Introduction), Meltzer, B. (Translator). Dover Publications, New York. Original publication, New York, Basic Books (1962). Electronic Reprint accessed at: jacqkrol.x10.mx on June 3, 2018 and accessed at: monoskop.org on June 3, 2018
31. Good, I.J.: A note on Richard's paradox. Mind **75**(299), 431 (1966)
32. Grim, P.: Philosophy of Mind: Brains, Consciousness and Thinking Machines. The Great Courses, The Teaching Company, Chantilly, Virginia (2008)
33. Hall, N.: David Lewis's metaphysics. In: Zalta, E.N. (ed.) The Stanford Encyclopedia of Philosophy, Winter edn. Stanford Metaphysics Laboratory, Stanford University. (2016). https://plato.stanford.edu/archives/win2016/entries/lewis-metaphysics/
34. Hansson, S.O.: Science and pseudo-science. In: Zalta, E.N. (ed.) The Stanford Encyclopedia of Philosophy, Summer edn. Stanford University, Stanford Metaphysics Laboratory (2017). https://plato.stanford.edu/archives/sum2017/entries/pseudo-science/
35. Harari, Y.: Homo Deus: A Brief History of Tomorrow. United Kingdom, Canada: Signal Books, an imprint of McClelland & Stewart, A Division of Penguin Random House Canada (2016)
36. Harris, S.: Free Will. Free Press, A Division of Simon and Schuster Inc., New York, London (2012)
37. Hatfield, G.: René descartes. In: Zalta, E.N. (ed.) Stanford Encyclopedia of Philosophy. Stanford University: Stanford Metaphysics Laboratory (2016). https://plato.stanford.edu/archives/sum2016/entries/descartes/
38. Hawking, S.: Kurt Gödel in God Created the Integers: Mathematical Breakthroughs that Changed History, p. 1265. Running Press Book Publishers, Philadelphia (2007a)
39. Hawking, S.: Kurt Gödel in God Created the Integers: Mathematical Breakthroughs that Changed History, p. 1258. Running Press Book Publishers, Philadelphia (2007b)
40. Hodges, A.: Alan turing. In: Zalta, E.N. (ed.) The Stanford Encyclopedia of Philosophy, Winter edn. Stanford: Metaphysics Research Laboratory, Stanford University (2013). https://plato.stanford.edu/archives/win2013/entries/turing/
41. Hofweber, T.: Logic and ontology. In: Zalta, E.N. (ed.) The Stanford Encyclopedia of Philosophy, Winter edn. Stanford Metaphysics Laboratory, Stanford University (2017). https://plato.stanford.edu/archives/win2017/entries/logic-ontology/
42. Humphreys, P.: Emergence: A Philosophical Account. Oxford University Press, New York (2016)
43. Irvine, A.D.: Principia mathematica. In: Zalta, E.N. (ed.) The Stanford Encyclopedia of Philosophy, Winter edn. Stanford Metaphysics Laboratory, Stanford University (2016). https://plato.stanford.edu/archives/win2016/entries/principia-mathematica/
44. Kauffman, S.: The Origins of Order: Self-Organization and Selection in Evolution. Oxford, Oxford University Press, New York (1993)
45. Kauffman, S.: Reinventing the Sacred: A New View of Science, Reason and Religion. Basic Books. A Member of the Perseus Group, New York (2008)
46. Kauffman, S.: Humanity in a Creative Universe. Oxford University Press, New York (2016)
47. Kennedy, J.: Kurt Gödel. In: Zalta, E.N. (ed.) The Stanford Encyclopedia of Philosophy, Winter edn. Stanford Metaphysics Research Lab, Stanford University (2016). https://plato.stanford.edu/archives/win2016/entries/goedel/

48. Kuhn, T.: The Structure of Scientific Revolutions. University of Chicago Press, Chicago (1962)
49. Kuhn, T.: The Structure of Scientific Revolutions, 50th Anniversary. University of Chicago Press, Chicago (2012)
50. Laughlin, R.B.: A Different Universe (Reinventing Physics from Bottom Down), pp. 31–32. Basic Books, A member of the Perseus Book Group, New York (2006)
51. Laughlin, R.B.: A Different Universe (Reinventing Physics from Bottom Down), p. 247. Basic Books, A Member of the Perseus Book Group, New York (2006b)
52. Libet, B.J.: Do we have free will. J. Conscious. Stud. **6**(8–9), 47–57 (1999)
53. Miller, J.H., Page, S.E.: Complex Adaptive Systems: An Introduction to Computational Models of Social Life. Princeton University Press, Princeton and Oxford (2007)
54. Mitchell, S.D.: Biological Complexity and Integrative Pluralism. Cambridge University Press, Cambridge (2003)
55. Mitchell, S.D.: Unsimple Truths: Science. Chicago, Chicago University Press, Complexity and Policy (2009)
56. Nagel, E., Newman, J.R. (1956, reprint 2000) Goedel's Proof, pp. 1668–1695. In: Newman, J. (ed) (1956, reprint 2000). The World of Mathematics, vol. 3: Parts viii-xvii. Dover Publishers Inc., Mineola, New York
57. Nagel, E.: The Structure of Science. Problems in the Logic of Explanation. Harcourt, Brace & World, Inc., New York (1961)
58. Nagel, E., Newman, J.: Gödel 's Proof. In: Hofstadter, D.R. (ed.) (Revised and Edited with a new Foreword by D.R. Hofstadter editor), pp. Foreward and Introduction. New York and London: New York University Press. (Reprinted from: Electronic EBook) (2001)
59. Nagel, T.: Mind and Cosmos: Why The Materialist Neo-Darwinian Conception of Nature is Almost Certainly False. Oxford University Press, Oxford (2012)
60. Nickles, T.: Scientific revolutions. In: Zalta, E.N. (ed.) The Stanford Encyclopedia of Philosophy, Winter edn. Stanford Metaphysics Laboratory. Stanford University (2017). https://plato.stanford.edu/archives/win2017/entries/scientific-revolutions/
61. Okasha, S.: Philosophy of Science: A Very Short Introduction. Oxford University Press, Oxford (2002)
62. Popper, K.: The Logic of Scientific Discovery. Hutchinson, London (1959)
63. Pylkkanen, P.: Mind, Matter and the Implicate Order, p. 2. Springer-Verlag Publishing, Heidelberg
64. Raatikainen, P.: Gödel's incompleteness theorems. In: Zalta, E.N. (ed.) The Stanford Encyclopedia of Philosophy, Spring edn. Stanford: Metaphysics Research Lab, Stanford University (2015). https://plato.stanford.edu/archives/spr2015/entries/goedel-incompleteness/
65. Raatikainen, P.: Gödel's incompleteness theorems, p. 2. In: Zalta, E.N. (ed.) The Stanford Encyclopedia of Philosophy, Spring edn. Stanford Metaphysics Research Lab, Stanford University (2015). https://plato.stanford.edu/archives/spr2015/entries/goedel-incompleteness/
66. Robinson, W.: Epiphenomenalism. In: Zalta, E.N. (ed.) Stanford Encyclopedia of Philosophy, Spring edn., p. 1 (2015a). http://plato.stanford.edu/archives/spr2015/entries/epiphenomenalism/
67. Robinson, W.: Epiphenomenalism. In: Zalta, E.N. (ed.) Stanford Encyclopedia of Philosophy, Spring edn. (2015b). http://plato.stanford.edu/archives/spr2015/entries/epiphenomenalism/
68. Schwartz, J.M., Begley, S.: The Mind & The Brain: Neuroplasticity and the Power of Mental Force. Harper Perennial Publishers, New York (2002)
69. Scott, A.: Physicalism, chaos and reductionism. In: Tuszynski, J.E. (ed.) The Emerging Physics of Consciousness, pp. 171–192. Springer, Heidelberg (2006)
70. Scott, A.C.: The Nonlinear Universe: Chaos, Emergence, Life, pp. 277–301. Springer, Heidelberg (2010)

71. Scott, R.: Letter to editor. Complexity **2**(6), 10 (1997)
72. Scott, J.R.: Series of papers in preparation. Presented as Poster presentation at 2018 ICCS conference in Boston Massachussets. Present paper accepted for proceedings of 2018 ICCS conference (2018)
73. Seager, W.: Natural Fabrications: Science, Emergence and Consciousness. Springer, Heidelberg (2012)
74. Seager, W.: Natural Fabrications: Science, Emergence and Consciousness, pp. 189–210. Springer, Heidelberg (2012)
75. Shapiro, Y., Scott, J.R.: Dynamical systems therapy (DST): complex adaptive system in psychiatry and psychotherapy. In: Mitleton-Kelly, E., Paraskevas, A., Day, C. (eds.) Handbook of Research Methods in Complexity Science, pp. 567–589. Edgar Elgar Publishing, Cheltenham (2018)
76. Siegel, D.J.: The Developing Mind: Toward a Neurobiology of Interpersonal Experience. The Guilford Press, New York (1999)
77. Siegel, D.J.: Pocket Guide to Interpersonal Neurobiology: An Integrated Handbook of the Mind. W.W Norton Inc., New York (2012)
78. Siegel, D.J.: Mind: A Journey to the Heart of Being Human. W.W. Norton & Company, New York (2016)
79. Steup, M.: Epistemology. In: Zalta, E.N. (ed.) The Stanford Encyclopedia of Philosophy, Fall edn. Stanford Metaphysics Laboratory, Stanford University (2017). https://plato.stanford.edu/archives/fall2017/entries/epistemology/
80. Stevens, A.: Private Myths: Dreams and Dreaming, pp. 297–318. Harvard University Press, Cambridge (1995)
81. Tabery, J.: Beyond Versus: The Struggle to Understand the Interaction of Nature and Nurture. The MIT Press, Cambridge (2014)
82. Unger, R.M., Smolin, L.: The Singular Universe and the Reality of Time: A Proposal in Natural Philosophy. Cambridge University Press, Cambridge (2015)
83. van Riel, R., Van Gulick, R.: Scientific reduction. In: Zalta, E.N. (ed.) The Stanford Encyclopedia of Philosophy, Winter edn. Stanford Metaphysics Laboratory, Stanford University (2016). https://plato.stanford.edu/archives/win2016/entries/scientific-reduction/
84. Wolfram, S.: A New Kind of Science, p. 846. Wolfram Media Inc., Champaign (2002)
85. Yanofsky, N.S.: Paradoxes, contradictions, and the limits of science. Am. Sci. **104**(3), 166–173 (2016)
86. Zach, R.: Hilbert's program. In: Zalta, E.N. (ed.) The Stanford Encyclopedia of Philosophy Spring Edition ed. Stanford Metaphysics Laboratory: Stanford University (2016). https://plato.stanford.edu/archives/spr2016/entries/hilbert-program/

Multi-scale Intent

Anne-Marie Grisogono(✉)

Flinders University, Adelaide, South Australia, Australia
annemarie.grisogono@flinders.edu.au

Abstract. A multi-scale framework of intents is described for purposeful agents as a tool to explore the dynamics of cooperation, collaboration, conflict and competition between humans, both as individuals and as groups. The framework spans multiple timescales over which intentions persist - from the most transient (the next action to be taken) to the most enduring (deeply held values to be upheld), with several implicit scales in between. It is argued that each agent's behaviour in a particular situation is shaped by its own intent framework and by certain aspects of its conceptual model of the situation (including itself). This model suggests a systematic approach to identifying the scope for cooperative behaviours between two or more agents, and moreover where and how the domain of cooperation could be enlarged, with applications to improving cooperation and reducing conflict in human groups.

Keywords: Multiscale intent · Cooperation · Intentional agents
Human conflict

1 Motivation

Interactions between humans are at the heart of most of our endeavours, from the local scale, face to face, give and take, sharing, squabbling, helping and scheming of small social groups, through to the largest scale of international and trans-national, global relations in commerce, production, security, culture and research. Whether the actors are institutions, or governments, armies or NGOs, insurgents or sports teams, the recurring themes are on the spectrum from peaceful and productive collaboration, cooperation, and co-existence through various levels of overt and covert competition, manipulation and exploitation, to outright violent conflict. The extent to which actors can achieve their goals is very dependent on the quality of interactions they have with other actors whose domains of action impinge on theirs.

It is important therefore to have a deep understanding of what factors shape the form these interactions take, and what avenues exist for modifying them in order to achieve more success and less harm.

All these interactions take place in complex environments, in the sense of consisting of multiple interacting agents, with diversity, autonomy, memory, intelligence and various levels of social structure, together with a similarly complex physical environment with diverse and distributed, but limited resources and affordances. It is well-known that one of the ubiquitous features of complex systems is their multi-scale structure [1], which is intimately linked with the notion of emergence. We see structure

A. J. Morales et al. (Eds.): ICCS 2018, SPCOM, pp. 53–61, 2018.
https://doi.org/10.1007/978-3-319-96661-8_5

at multiple scales, and the associated emergent phenomena at each scale, not only in human organisations and enterprises, but everywhere in living systems, and in the complex systems we create. The scales form a natural partitioning for our attention, because the entities and processes we observe at a particular scale seem coherent and 'sufficient' for description or explanation, but the scare quotes are there because of course it is not in general strictly true that one can afford to pay attention only to what is happening at one scale. There are important and deep relationships between the phenomena at different scales, and it is often the case that a problem perceived at one scale is a symptom or consequence of one or more other conditions at other scales, and unless these are identified and addressed, the perceived problem is unable to be resolved.

This is a fruitful line of thinking to apply to the problem of understanding and fostering the potentials for cooperative engagement between actors.

2 A Single Actor's Multi-scale Intent

Let us first consider a single actor (whether an individual, an organization, a state or an alliance) engaging in a complex situation.

What we mean by 'complex' here is that the underlying dynamics of the situation are interlinked and nonlinear, so that the consequences of an action may play out through multiple interacting pathways and similarly, any event or property of the situation arises as a result of many interacting causal and influence paths. We assume that the actor engages in the situation with some intention – i.e. the actor will wish to influence the development of the situation in some way.

In order to make progress in what is already a very complex scenario, we make some simplifying assumptions. Fully acknowledging that the fields of psychology, cognition, anthropology and social science will paint vastly more intricate and detailed pictures [2], we posit a simple, but multi-scale, 'cartoon' of the actor's internal conceptual structures, loosely based on a hierarchy of ends, ways and means, as shown in Table 1 below.

In this rudimentary model, actors' choices and behaviors are based on a bedrock of relatively enduring core values, but successive decomposition into more detailed preferences and choices in particular situations are strongly influenced by their conceptual understanding of the dynamics of those situations at the corresponding relevant scales:

- What actors judge to be **successful or failed outcomes** in the situation are derived from their understanding of what outcomes are possible in the situation or overall environment and how those outcomes stack up against their **core values**.
- To get beyond simply observing and judging the situation implies a **strategy**, that is a concept of how to deal with the environment as a whole at a given time and
- a concept of how to influence a specific evolving situation within that environment towards a desired 'end-state' – the actor's ***stratagem***.
- These in turn give rise to a **course of action** – a set of proposed **objectives** and a concerted approach, and

Table 1. An Actor's Multi-scale Intent Hierarchy. (*detailed discussion in text below*)

<table>
<tr><th>How frequently changed</th><th>What actor wants or intends to achieve
(ends --- means)</th><th>Actor's concept of how to achieve Intention
(ways)</th><th>Dependence on actor's understanding of context</th><th colspan="2">needed contextual appreciation:
scope | resolution</th></tr>
<tr><td rowspan="9">Enduring
↓
as context evolves and understanding of it develops
↓
as situation evolves and understanding of it develops
↓
moment by moment</td><td>Defend core values</td><td></td><td>none</td><td colspan="2">not applicable</td></tr>
<tr><td rowspan="2">Protect/advance interests
Achieve desired overall end-state
Metric: success/ failure</td><td rowspan="2">strategy</td><td>high</td><td>very wide</td><td>low</td></tr>
<tr><td>very high</td><td>very wide</td><td>low</td></tr>
<tr><td rowspan="2">Desired operational end-state
Metrics: indicators/proxies</td><td>stratagem</td><td>very high</td><td>wide</td><td>mid</td></tr>
<tr><td>course of action</td><td>high</td><td>restricted</td><td>high</td></tr>
<tr><td rowspan="2">Desired outcomes / effects / objectives</td><td rowspan="2">plan</td><td>high</td><td>restricted</td><td>high</td></tr>
<tr><td>medium</td><td>narrow</td><td>higher</td></tr>
<tr><td rowspan="2">tasks</td><td rowspan="2">procedures</td><td>low</td><td>specific aspects</td><td>very high</td></tr>
<tr><td>low</td><td>local</td><td>very high</td></tr>
</table>

- a detailed if flexible **plan** with specific tasks to achieve the desired effects and immediate and long term end-states.

The defining scale factor in this model is time in the leftmost column, with the corresponding scale of intent in the adjoining column to the right. Thus, time ranges from the immediate (what the actor intends doing next: laying a brick, delivering a message, and so on, which obviously changes from one moment to the next, through to the very long term intentions which are implicit in the actor's espoused values: if they value kinship, or adherence to a religious code then these are enduring aspects of their overall posture in the world, and will tend to be very resistant to attempts to change them. They will also colour the relevant choices the actor makes at the other scales. In between these two extremes we posit intermediate scales which are the links between enduring values and observable actions. Taking a cue from military planning [3], we identify intermediate levels. Short-term tasks consist of procedures which the actions taken contribute to achieving. Mid-term plans are resolved into tasks, and the desired outcomes of the plans are waypoints on longer courses of action, so observing their achievement are proxy measures for eventual success. The courses of action in turn combine to form the stratagem for influencing what they care about in the situation, and the stratagem is an instance of their overall longer-term strategy for engaging with the situation.

Of course this can immediately be criticized as being not at all how real people develop their intentions and choose their actions. Nevertheless we argue in the following paragraphs, first that it is feasible to apply, and secondly, that it is a useful framework.

Even if the actor has not thought through a strategy and its component courses of action, there will still be some kind of rationale for what they are doing and it can be discovered (if they are willing to be truthful) by asking why they are performing the current action. The answer is likely to be of the form *"..in order to… achieve X*", which gives an indication of the intent at the scale above the observed action. The next question is why they want to achieve X and once again we may hear *".. in order to … achieve Y*" – which indicates an element of intent at the next scale. The process terminates when the reply is along the lines of *".. just because that's what I want*" which links directly to deeply held values.

It may not be a very coherent or persuasive rationale, and it need not have the number of levels described here, it also may not be truthful, either through deliberate concealment or unconsciously through lack of self-awareness of their own cognitive biases and predispositions [4] – all these possibilities can be factored in - but it can still be very useful to elicit it as we will discuss. In the event that the actor is honestly unable to give any explanation of their actions – they may well be picking actions on a whim or randomly because they have no particular purpose in relation to the situation, or because they are confused, and this may also be useful information.

To explore the utility of this framework we start with the observation that the actor's understanding of the dynamics of the environment, how a particular situation may play out, what actions might shape outcomes in a particular direction, and what actions they are capable of implementing, are critical factors in how they arrive at their intents at each scale. So a change in their understandings (through either learning that the situation has changed, or learning that an aspect of their previous understanding was incorrect or incomplete), and/or a change in their understanding of their own capabilities, should result in a change to intents at the appropriate scale.

Ideally, such understandings also include an appreciation of the wide variety of complex factors that might drive the situation towards success or failure. So as indicated in the rightmost columns of Table 1, the *scope* of understanding required for the development of strategy is very wide, with that for a stratagem only slightly less so, but in order to see the wood, there should not be too much detail in the trees – i.e. the degree of resolution needed is very low.

As we move down the table, the scope narrows as attention is focused on particular aspects of the situation and on more specific actions and their consequences, but the level of detail required increases correspondingly. At the lowest level, the scope may narrow to a small focal area, with actions adapted from standard procedures but with a high degree of resolution required commensurate with the task at hand.

The relative impact of the quality of the decision-makers' understanding of the relevant aspects of the situation (including of their own capabilities) on the intent that they form at that scale, is colour-coded in the fourth column of Table 1. Thus at the scales of overall strategy, and specific stratagems and courses of action, the decisions made are highly dependent of the quality of understanding, but much less so at the scales of procedures, while values are principles that should hold in any situation.

While this crude model can easily be critiqued from many angles, we feel that it has some merit in enabling us to explore the interactions between and among the actors who may be seeking to influence the same situation. We are interested in the conditions under which they will be willing to cooperate, at what level they will be willing to cooperate, how robust that cooperation may be when conditions change, and most importantly, what can be done to foster the growth of such cooperation and to increase its robustness.

3 Emergence of Cooperation Between Two Actors

A key feature of this construct is the attempt to make explicit the impact of the actors' understanding of the situation and its dynamics on the choices they make.

For example, two actors with the same core values but somewhat differing interests may arrive at quite different appreciations of the situation and therefore end up with very different measures of success and failure.

Similarly, agreement on a desired end-state does not guarantee that the actors will agree on the stratagems, courses of action or detailed plans for achieving that end-state because the actions to shape that end-state will have different spin-off effects that will affect their individual interests differently. Thus, there will be different opinions about how the system dynamics operate that will be exacerbated if there is no agreement on an underlying conceptual framework.

When we are dealing with a complex situation, these differences can have major impacts, since we know that the complexity guarantees that every perspective will be limited, flawed and incomplete – in other words there is no single complete correct view which everyone can be brought to agreement on.

Let us take the simplest case of two actors, each with their own interests and hierarchy of situation understanding, intentions and plans as in Table 1. A hypothetical third party with a helicopter view could compare and note the extent of agreement or disagreement between their corresponding levels of situational understanding, and the extent of potential synergy or conflict between what they currently intend to do.

What each actor actually perceives of the other's interests and hierarchy of situation understanding, intentions and plans is of course open to error of various kinds, but an actor who believes that there is potential synergy between his own actions for example, and those of another, will be motivated to cooperate if he believes that doing so would lead to a better overall probability of success as he defines it. Willingness to cooperate implies that each is prepared to modify his own thinking, planning and actions in order to realize the potential synergy.

But "success" is in the eye of the beholder. In a relationship between two complex actors, it is unlikely that there will not be at least nuanced differences in the objectives or effects sought, on the details and execution of even an agreed upon course of action, in judgments as to the correct indicators of success, and as a result, differing conclusions as to potential synergies. So, where can our actors go from here?

Let us suppose both believe there is potential synergy between proposed courses of action – even if for different reasons. That means that there is a positive gradient in both their 'fitness landscapes' that they can climb, so provided they both judge the benefits worthy of the costs, cooperation should follow. However, this could be a very fragile cooperation. As the situation evolves – and it will – and as their respective situational understandings develop, their intents must also change. What is the chance that the originally perceived synergy – the basis for cooperation, will survive?

Our conjecture is that synergies of accident (i.e. for different or opposing reasons) are most likely to dissolve when changes occur, while synergies based in enduring core values, interests and a common outlook and grand strategy and therefore a more consistent context and conceptual base will be more likely to adapt in concert, and thus be robust as the situation develops. This is what we mean by coherence in this context. If this is correct, and we wish to foster cooperation and coherence, then we must do more than compare proposed courses of action looking for apparent synergies; we must also examine and understand the other actor's values, interests, desired end-states and proxies for success as outlined in Table 1.

Doing so allows us to identify potential sources of divergence, whether primarily in differences of understanding with regard to one or another aspect of this specific situation, or in more fundamental and intransigent differences of values and interests.

If it is the former, and there is a shared core foundation, then collaborating in the development of situational understanding may lead to convergence and thus offer a possible route to adaptively increase coherence. If values and interests diverge significantly and there is little common ground on a desired end-state then there is no strong basis for developing coherence, although limited cooperation may still be possible wherever objectives are not mutually exclusive and the two actors can agree on some specific tasks that need to be performed.

As the above underlines, the position is unlikely to be black and white. There will be domains of agreement, domains in which each focuses on different but not inconsistent aspects, and some areas of possibly overt disagreement. Clarifying this space will permit the actors to delineate where it is possible to develop robust coherence through synergistic planning based on collaboratively developed understanding, where deconflicted complementary plans might be separately developed, or where it is best to simply stay out of each other's path.

We now pick up again the question of how and what an actor might be willing to change in order to realize a potential benefit. Suppose A and B discuss their interests, objectives, desired effects and end states, measures and indicators/proxies of success, their ways (strategy, stratagems and proposed courses of action) and means (capabilities, plans, tasks, and ways of working) in order to determine whether and how they might collaborate. In doing so, each weighs up what he may need to, or be willing to, change, what the other is willing to change and what the benefit impact of such changes would be on the achievement of his ends.

A's benefit in this context could mean that A's ends are achieved with less of A's effort, or faster, or more comprehensively. Of course if neither of them makes any change, then any benefit one gets from the other's actions is not the result of cooperation but of coincidence. The construct of Table 1 also helps clarify what kinds of changes are possible, and how a given change might result in benefit distributions ranging from totally one-sided to equal. As a result of interacting with B, A might decide to change his intents in the situation (e.g. because he has come to understand better what is possible, or what consequences might flow from his previous intent), or he may retain his intents but change his chosen course of action (e.g. because he learns that it is no longer likely to achieve his desired outcome, or because better possibilities arise from cooperating with B – so B is going to be making changes too).

In general, a change in A's ends (i.e. objectives, effects and end-states) should precipitate a corresponding change in specific aspects of A's ways and means so as to realign them with his new ends, while a change that only enters A's hierarchy at a lower level (for example a change of A's course of action) will affect his downstream plans and tasks but not his intentions with respect to his upstream stratagem and success and failure measures, although the outcomes of implementing the change will of course affect what is achieved against those intentions. So A may be holding to all his previous ends and ways, but the cooperation with B may make better means available for implementing them, thus improving his likely achievements against his ends.

Overall, a given change may enhance a shared end (in which case both benefit), an end valued by one but not the other (in which case the benefit is asymmetric) or more than one end – for example one favored by A and another favored by B, in which case various benefit distributions may be reached.

One might be tempted to claim that unless there was benefit to both sides it would not be cooperation, however there are finer distinctions to be made here – the benefit from helping a less mature partner to improve their effectiveness in pursuing ends that one does not care about, might rather accrue over longer timescales and wider scopes than presently considered. Benefit might also derive not from the direct impact, but rather from indirect effects of those changes – for example through their influence on the perceptions and behaviors of third parties.

However, in general, it is reasonable to expect that cooperation usually leads to shared benefits, at the very least on the strength of enhanced outcomes against those ends to which both are willing to subscribe. If there were no such agreed ends, one might expect that any cooperation would be very fragile and accidental.

Clearly, the possible interactions between two actors with dynamic hierarchies of situational understandings, intents and capabilities, are already complex enough, and our discussion is not exhaustive, but it suffices to illustrate the richness of the two-actor interaction space. To explore this space further, consider the average distributions of changes and benefits that result from their repeated interactions over some time. Let one axis of the space denote who does the changing (say A does all the changing at one end, vice versa at the other, and equal amounts in the middle) and the other axis show who benefits from the change (similarly: from A gets all the benefit and B gets none, to vice versa, and equal benefits in the middle). In brief, the midpoint of this space represents perfectly symmetric cooperation, the ends of one diagonal represent one exploiting or manipulating the other, while the ends of the other diagonal represent independence.

As usual though, it is not the extremes that are most relevant but the intermediate ranges. Moving across the trade space such that A's objectives always benefit more than B's from changes brought about by their interactions, we note that at one end, it benefits mainly from its own changes – in other words it is adjusting its ways and means to better achieve its own ends, and at the other mainly from persuading B to make changes to its ways and means to align better with A's ends.

Another dimension to be considered is the magnitudes of changes and benefits. Obviously a large share of a small benefit might not be as attractive as a smaller share of a much larger benefit. So this means we should also consider what the potential impact of the cooperation might be – where are the greatest payoffs to be found?

Interestingly, and perhaps counter-intuitively, we propose that they are not necessarily in the middle "equal" zone.

First we note that the potential benefits of cooperation between two actors who have similar objectives fall into two broad categories: *complementary cooperation* where the benefits stem from the relevant different capabilities that each brings, and *supplementary cooperation* where the benefits stem from the additional similar capabilities that each brings. Now, interactions in the "equal" zone imply high congruity between the intents of the two actors. Of course high congruity of intents can arise in very different actors, but to the extent that the congruity stems from more extensive similarities, for instance in their previous histories, in their knowledge base, in their relationships with other significant actors, and in their abilities and ways of working, then the potential impact of their cooperation is likely to be more limited than it would be with actors that add more to each other's diversity. For example, cooperation from an actor that had closer relationships with certain target groups in the situation might be much more valuable for advancing one's objectives, than one that could add capacity to an existing strength.

On the other hand, the more different two actors are, the less likely they are to find easy agreement on each element of Table 1, so achieving cooperation is going to be more difficult, require more investment of time and effort by both, and also require more effort to maintain it as the situation develops. However the return on that effort may be very significant for each actor through complementary actions by the other that shape the environment in ways that render his own actions more feasible and effective. Our very general conclusion is that the more similar two actors are, the easier it may be to achieve cooperation, but the less value may be gained by each from doing so, while the more different they are, the more potentially valuable, and the more potentially difficult it will be. This suggests that we need a range of strategies for engendering cooperation rather than a unitary approach. We postulate that if potential partners A and B are similar in that they have strong overlap in their interests and concepts of success and failure for the situation, they will easily operate in the middle (green) region since any changes that improve outcomes against a success measure or a proxy for one, will be valued by both. What varies left to right across the zone is the relative maturity and competence of A and B. In this case, the focus is on using one's strengths to help the other's weakness. If the weakness of actor B is in his situational understanding then A's cooperating with him in its development will not only strengthen B's ability to act effectively in pursuit of his own ends, but more importantly, will help to bring about closer alignment of his actions with A's. So this cooperation strategy will readily grow coherence. Power balance is an issue to the extents that each also aspires to ends that are not supported by the other, but of course this will always be the case to some degree with real actors. The temptation for the actor who is more powerful is to put pressure on the weaker to divert some of their effort to supporting the stronger partner's objectives at the expense of their own. The current return on that strategy may well justify the foregoing of possible future benefits from cooperation. For potential partners that have strong rivalry and clashes of values, the situation is a little more complex. They might be maneuvering along the power see-saw diagonal especially if the focus of the interaction is on their differences with respect to their values and success and failure goals, or fundamental differences in their perspectives on the drivers of the situation's dynamics. Nevertheless, they might also have a lot to gain from carefully negotiated cooperation in defined domains.

The challenge for the actors in this kind of relationship is to defuse the power struggle and to step outside of rigid traditional roles and views in order to make space for finding and nurturing the mutual benefit domain, which could be very significant indeed, simply because they are so different. Such a cooperative relationship is not likely to ever become coherent since the underlying agendas on each side are so different, but steps could be taken to make it as robust as possible.

A fuller and more thoughtful discussion of the many issues alluded to here is certainly needed, but beyond the scope of the present paper.

References

1. For discussion of the multi-scale hierarchical structure of complex systems see for example Bar-Yam, Y.: Multiscale variety in complex systems. Complexity, **9**(4), 37–45, March–April 2004
2. Some well-known models of human decision–making: Anderson, J.R. (1991). The place of cognitive architectures in a rational analysis. In: Van Len, K. (Ed.): Architectures for Intelligence. Hillsdale, NJ: Erlbaum; Klein, G.A., Orasanu, J., Calderwood, R., Zsambok, C. E. (Eds.) (1993). Decision making in action: Models and methods. Westport, CT, US: Ablex Publishing. However they do not address the multi-scale structures that link actions to values and to the decision-maker's conceptual models at multiple scopes and resolutions
3. For example: Army Doctrine Publication (ADP) 5-0 (FM 5-0), The Operations Process. Department of the Army Washington, DC. Accessed 17 May 2012. https://militarydecisiongr.files.wordpress.com/2015/05/a-guide-to-the-mdmp.pdf
4. Kahnemann, D.: Thinking, Fast and Slow. Penguin, 2011; Arnott, D. Cognitive biases and decision support systems development: a design science approach. Inf. Syst. J., 55–78, 1 Jan 2006

The Laws of Complexity and Self-organization: A Framework for Understanding Neoplasia

Nat Pernick(✉)

30100 Telegraph Rd, Suite 408, Bingham Farms, MI 48025, USA
NatPernick@gmail.com

Abstract. Background: Current biologic research is based on reductionism, through which organisms and cells are merely combinations of simpler systems. However this approach has failed to substantially reduce cancer-related deaths. Complexity theory suggests that emergent properties, based on unpredictable, nonlinear interactions between the parts, are important in understanding fundamental features of systems with large numbers of independent agents, such as living systems.

Methods and Findings: The laws of complexity and self-organization are summarized and applied to neoplasia:

1. In life, as in other complex systems, the whole is greater than the sum of the parts.
2. There is an inherent inability to predict the future of complex systems.
3. Life emerges from non-life when the diversity of a closed system of biomolecules exceeds a threshold of complexity.
4. Much of the order in organisms is due to generic network properties.
5. Numerous biologic pressures push cellular pathways towards disorder.
6. Organisms resist common pressures towards disorder through multiple layers of redundant controls, many related to cell division.
7. Neoplasia arises due to failure in these controls, with histologic and molecular characteristics related to the cell of origin, the nature of the biologic pressures and the individual's germline configuration.

Conclusions: Cells maintain order by redundant control features that resist inherent biologic pressures towards disorder. Neoplasia is due to the accumulation of changes that undermine these controls. Studying neoplasia within this context may generate new therapeutic approaches by focusing on the underlying pressures on cellular networks.

An expanded version of this paper is available at http://natpernick.com/TheLawsJune2017.pdf.

1 Introduction

1.1 The War on Cancer

On 23 December 1971, President Richard M. Nixon signed the National Cancer Act of 1971, generally viewed as beginning the "war on cancer" in the United States [1]. Fifteen years later, Bailar and Smith concluded that "we are losing the war against

A. J. Morales et al. (Eds.): ICCS 2018, SPCOM, pp. 62–70, 2018.
https://doi.org/10.1007/978-3-319-96661-8_6

cancer, notwithstanding progress against several uncommon forms of the disease, improvements in palliation, and extension of the productive years of life" [2]. Recent data indicate that the 5-year relative survival rate has increased from 49% in 1975–1977 to 69% in 2005–2011 [3]. However, although the U.S. National Cancer Institute has spent over $100 billion on this effort [4], progress has been limited in reducing mortality from common, advanced carcinomas of the lung, colon, breast, and pancreas, and overall U.S. cancer deaths are projected to rise to 609,640 in 2018.

1.2 Reductionism: The Current Approach to Biology

Current research efforts in biology are based on the reductionist approach, summarized as "the whole is equal to the sum of its parts". This "gold standard" for learning about the world is based on the works of Descartes, Galileo, Newton and LaPlace, postulating that the workings of our mind and body and all matter in the universe unfold under the same set of fundamental laws [5]. With this approach, cells can theoretically be completely understood by analyzing all components and the connections between them, which are assumed to be additive and linear [6, 7]. Under this view, diseases are studied by finding and understanding defective genes, proteins, or other biomolecules in a cell, tissue, or organ. For example, follicular lymphoma is due to the t(14;18) (q32; q21) translocation, present in 80–90% of tumors, which brings the *bcl2* proto-oncogene under the transcriptional influence of the immunoglobulin heavy chain gene, leading to overexpression of the Bcl2 protein, which inhibits apoptosis. This inhibition allows additional genetic mutations to accumulate, which leads to neoplasia [8]. But this reductionist model does not explain the myriad network changes facilitating the translocation or the web of network changes it induces.

The goals of this paper are to discuss how complexity theory may relate to neoplasia, to explain to the pathology community why the reductionist model is inadequate and to suggest that effective cancer research should incorporate the laws of complexity and self-organization.

1.3 Complexity: Variability that Is not Predictable

Complexity refers to systems with large numbers of independent agents with a high and variable degree of connectivity [9]. Complex systems exhibit many nontraditional properties [10]. First, they have variable behavior that obeys the laws of physics, but cannot be reliably predicted by reproducible experiments [11]. Behavior also varies due to self-organized criticality, a dynamic process that drives large extended systems to a network state that is poised at criticality, analogized to a sand pile created by dropping individual sand grains [11]. Small avalanches may be predictable, but the overall behavior of the sand pile is best described by catastrophic, not gradual, changes.

Second, complex systems possess a robustness that makes them resistant to significant changes. The maintenance of cellular phenotypes and stability in physiologic processes has been attributed to "attractors" associated with a complex gene regulatory network, which maintains and reestablishes specific gene expression patterns, even after large perturbations [12].

Third, complex systems possess emergence, an organizational, bottom-up property, due to agents that spontaneously self-organize without any oversight or planning [9] (p. 11). Larger entities arise through interactions among simpler entities and possess properties or exhibit features not found or even thought possible from the simpler entities and that require fundamental research to understand. In biological systems, self-organization has been described as a process in which global patterns emerge solely from numerous lower level interactions, even though the rules specifying interactions are executed using only local information [13]. Neoplasia cannot be well understood without knowledge of emergence.

Fourth, similar appearing behavior and features may be due to markedly different inputs. In colon carcinoma, alterations to dissimilar molecular pathways may produce morphologically similar tumors [14].

2 The Laws of Complexity and Self-organization

The Laws of Complexity and Self-Organization relevant to neoplasia are:

1. In life, as in other complex systems, the whole is greater than the sum of the parts.
2. There is an inherent inability to predict the future of complex systems.
3. Life emerges from non-life when the diversity of a closed system of biomolecules exceeds a threshold of complexity.
4. Much of the order in organisms is due to generic network properties.
5. Numerous biologic pressures push cellular pathways towards disorder.
6. Organisms resist common pressures towards disorder through multiple layers of redundant controls, many related to cell division.
7. Neoplasia arises due to failure in these controls, with histologic and molecular characteristics related to the cell of origin, the nature of the biologic pressures and the individual's germline configuration.

2.1 In Life, as in Other Complex Systems, the Whole Is Greater Than the Sum of the Parts

The reductionist approach is inadequate for understanding living systems and diseases such as cancer; biology cannot be reduced to physics alone [5]. In living systems, the interactions between molecules create life. Individually, the molecules can be considered as "dead." Collectively, they develop emergent properties, the missing features that make the whole greater than the sum of its parts [15]. Mitosis is an emergent property with obvious importance in neoplasia. Various molecules engage in linked processes whose end result cannot be predicted even by examining a large subset of the processes. As Kauffman notes, "it is a closure of work tasks that propagates its own organization of processes" [5] (p. 94).

2.2 There Is an Inherent Inability to Predict the Future of Complex Systems

In 1814, Laplace claimed that one could determine the entire future and past of all particles in the universe and their motions if supplied with their instantaneous positions and velocities [16]. However, the ability to predict planetary motion or the tides does not extend to complex systems, for several reasons.

First, the chaotic nature of complex systems precludes predictability. Chaotic properties are characterized by nonlinear equations, which are exquisitely sensitive to initial conditions. Lorenz found that his computer model of the weather experienced exponential divergence when he reran it substituting the Fig. 0.506 for 0.506127 [17]. This inability to predict the future of systems that are well understood is an inherent property of the nonlinear world in which we live. Second, emergent properties are not predictable. In neoplasia, we can document the presence or absence of specific mutations but cannot precisely predict their impact. Third, the function of molecules may be dependent on evolutionary pressures, which themselves cannot be predicted [18]. Selection may favor individuals heterozygous for the human sickle cell mutation at codon 6 of the *beta* gene, but only in geographic areas where falciparum malaria is endemic, where this mutation protects erythrocytes from infection [19]. However, we cannot predict the impact of this particular mutation on survival in the local environment without knowing the evolutionary pressures of all other human molecules and how they reinforce or counteract each other.

2.3 Life Emerges from Non-life When the Diversity of a Closed System of Biomolecules Exceeds a Threshold of Complexity

According to Kauffman, life is the emergent collective property of a modestly complex mix of biomolecules (DNA, RNA, proteins, and others) which catalyze each other's formation [20] (Chapter 7). Individually, each molecule is relatively inert. However, with a large enough collection of molecules of sufficient complexity, confined to a small space to promote interaction, a self-sustaining web of reactions may form that can reproduce and evolve [21, 22].

This model of the origin of life may explain why free living cells have an apparent minimal complexity. *Mycoplasma mycoides JCVI-syn1.0* [23] and *M. genitalium* [24] are the smallest known genomes that constitute a cell, with 473 to 482 protein-encoding genes, a large number for the simplest organism. A collection of fewer genes would apparently lack the complexity to create a self-staining network.

2.4 Much of the Order in Organisms Is Due to Generic Network Properties

Each cell coordinates the activities of 20,000 genes and their products [25]. Activities as complex as mitosis occur through spontaneous interaction of biomolecules without external oversight. To obtain a deeper understanding of cancer, we need to better understand how order arises in cells. The traditional view is that the sole source of order in organisms is natural selection as described by Darwin. An alternative view is that order is an expected emergent property of molecular networks, based on structural properties of networks not dependent on details of the particular molecules [20].

Genes, RNA, and proteins form a complex parallel processing network in which molecules are connected to other molecules and control their activation. Theoretically a cell with 20,000 types of gene products, one copy of each and two possible properties for each gene product would have a state cycle of length $2^{20,000}$, or approximately $10^{6,000}$. However, a state cycle this large does not happen due to the surprising finding that if each gene product is regulated by at most two inputs, the median length of the state cycle is only the square root of the number of gene products, or 141 if N is 20,000 [20, 26]. This network property creates inherent stability even in networks with large numbers of gene products, as the cell network is localized to a very small percentage of its possible state space. In addition, stability is promoted when genes are regulated by "canalyzing" Boolean functions [27, 28], which means that one input can completely determine the property of the gene.

The ability of cells to maintain stable phenotypic states is due to the settling down of a gene regulatory network into attractors [29], what Kauffman terms "order for free". Mutations can change functional connections but usually do not greatly change the stability of the network due to these order inducing properties.

2.5 Numerous Biologic Pressures Push Cellular Pathways Towards Disorder

Tension exists in living systems between order and disorder, a result of the tradeoffs inherent to achieve compromise between conflicting interests [30]. Order is required for proper functioning of cells, tissues, and organs. Yet network flexibility is required for development, inflammation, and adapting to numerous environments. Neoplasia subverts the physiologic mechanisms that provide this network flexibility and prevents reversion to an ordered state [31]. To understand neoplasia better, it is important to understand how physiologic disorder arises, how cells manage it, and how neoplasia disrupts it.

First, creating an autocatalytic network promotes disorder, as it produces an increasing number of new molecules, which catalyze further reactions. Second, natural selection disfavors rigid order in living systems, which would doom species amidst environmental shifts [32]. Third, the ability of living systems to maintain viability after mutational changes demonstrates an inherit flexibility not present in a completely ordered regime. Kauffman believes that organisms maintain a position between order and disorder that he terms the "edge of chaos," an evolution-derived compromise between order and surprise that may be optimal to coordinate complex activities and to evolve further [33] (p. 86) [34]. Finally, physiologic biologic pressures promote disorder. Infections, infestations, autoantigens, inflammation, and hormone expression, alone and particularly in combinations, push some cells into an active cell cycle, a less stable state, and eventually into neoplasia.

2.6 Organisms Resist Common Pressures Towards Disorder Through Multiple Layers of Redundant Controls, Many Related to Cell Division

Organisms have multiple layers of redundant controls that resist these pressures towards disorder. First, based on interactions between the components, a large “frozen” component forms, whose state does not easily change over time, even as the states of other molecules change [20, 35]. Second, cellular membranes act as “border controls” to limit the entry of novel molecules that might create new reactions or alter existing ones and to compartmentalize existing molecules to limit unexpected reactions. Third, cells have robust processes to limit errors during cell division, such as DNA repair [36], which dramatically reduce transcription error rates [37]. Fourth, cells have several mechanisms to respond to injury or DNA damage, which might eventually alter proteins and pathways, including apoptosis, cycle arrest, autophagy, or protein synthesis shutoff [38]. Fifth, key cellular processes have numerous controls that tightly regulate their activity, such as delay of cell cycle progression during mitosis in the presence of DNA or spindle damage [39, 40]. Finally, the immune system is a final supervisory system of error correction by destroying cells with disordered properties [41]. Their importance is suggested by the association of immunosuppression with a markedly elevated risk of malignancy [42].

2.7 Neoplasia Arises Due to Failure in These Controls, with Histologic and Molecular Characteristics Related to the Cell of Origin, the Nature of the Biologic Pressures, and the Individual’s Germline Configuration

The laws of complexity and self-organization provide a framework to better understand neoplasia, which is required for optimal cancer treatment. Cells are end product of networks with emergent features whose ultimate impact often cannot be predicted (laws 1–3). Although these networks possess a great deal of stability (law 4), they are under constant pressure to breach the control mechanisms that maintain order (law 5). Only the presence of multiple redundant controls at various levels leads to adequate order and function (law 6), consistent with the multiple-hit theory of neoplasia [43, 44].

Cell of Origin. A neoplasm’s characteristics are related to the network state of the cell of origin, the nature of the biologic pressure, and the germline configuration. The cell’s network state determines response to cellular pressures. For example, the t(14;18) translocation is apparently only found in B lymphocytes [45] and is due to an illegitimate V(D)J recombination, an activity restricted to B cells [46].

Nature of Biologic Pressures. We have proposed that an alternative classification to morphology or molecular changes characterizes neoplasia by the nature of the biologic pressures [47]. For example, gastric MALT lymphomas are caused not by mutations, but by antigen-driven lymphoproliferation.

Germline Configuration. The nature of the neoplasia is affected by the germline configuration, including familiar cancer syndromes [48] as well as more subtle variations in networks affecting any of the numerous control factors described above.

3 Summary

The original contributions of this paper are (a) proposing that the failures of the War on Cancer are due to medicine's rigid adherence to reductionism; (b) summarizing complexity and self-organization as they relate to neoplasia; (c) proposing that scientists study chronic pressures that disturb physiologic networks leading to neoplasia; and (d) suggesting that treatments which reverse these pressures or alter networks towards less lethal pathways may be useful.

Acknowledgments. The author thanks Christine Billecke, PhD, for her excellent editorial assistance in preparing this manuscript.

References

1. Institute NC: National Cancer Act of 1971. National Cancer Institute (2016). https://www.cancer.gov/about-nci/legislative/history/national-cancer-act-1971. Accessed 27 May 2018
2. Bailar 3rd, J.C., Smith, E.M.: Progress against cancer? N. Engl. J. Med. **314**(19), 1226–1232 (1986)
3. Society AC: Cancer Facts & Figures 2016. American Cancer Society, Atlanta (2016)
4. Kolata, G.: Grant system leads cancer researchers to play it safe. N. Y. Times (2009). http://www.nytimes.com/2009/06/28/health/research/28cancer.html. Accessed 21 Nov 2016
5. Kauffman, S.A.: Reinventing the Sacred: A New View of Science, Reason and Religion. Basic Books, New York (2008)
6. Mazzocchi, F.: Complexity in biology: exceeding the limits of reductionism and determinism using complexity theory. EMBO Rep. **9**(1), 10–14 (2008)
7. Van Regenmortel, M.H.: Reductionism and complexity in molecular biology: scientists now have the tools to unravel biological and overcome the limitations of reductionism. EMBO Rep. **5**(11), 1016–1020 (2004)
8. Ott, G., Rosenwald, A.: Molecular pathogenesis of follicular lymphoma. Haematologica **93** (12), 1773–1776 (2008)
9. Waldrop, M.M.: Complexity: the emerging science at the edge of order and chaos. Simon & Schuster, New York (1992)
10. Rickles, D., Hawe, P., Shiell, A.: A simple guide to chaos and complexity. J. Epidemiol. Community Health **61**(11), 933–937 (2007)
11. Bak, P.: How Nature Works: The Science of Self-organized Criticality. Copernicus, New York (1996)
12. Huang, S., Ernberg, I., Kauffman, S.: Cancer attractors: a systems view of tumors from a gene network dynamics and developmental perspective. Semin. Cell Dev. Biol. **20**(7), 869–876 (2009)
13. Camazine, S.: Self-organization in Biological Systems (Princeton Studies in Complexity). Princeton University Press, Princeton (2001)
14. Colussi, D., Brandi, G., Bazzoli, F., Ricciardiello, L.: Molecular pathways involved in colorectal cancer: implications for disease behavior and prevention. Int. J. Mol. Sci. **14**(8), 16365–16385 (2013)
15. Corning, P.A.: The re-emergence of "emergence": a venerable concept in search of a theory. Complexity **7**(6), 18–30 (2002)

16. Hawking, S.: Does God play dice? (1999). http://www.hawking.org.uk/does-god-play-dice.html. Accessed 23 Nov 2016
17. Dizikes, P.: When the Butterfly Effect Took Flight. MIT (2011). https://www.technologyreview.com/s/422809/when-the-butterfly-effect-took-flight/. Accessed 23 Nov 2016
18. Allison, A.C.: Polymorphism and natural selection in human populations. Cold Spring Harb. Symp. Quant. Biol. **29**, 137–149 (1964)
19. Sabeti, P.: Natural selection: uncovering mechanisms of evolutionary adaptation to infectious disease. Nat. Educ. **1**(1), 13 (2008)
20. Kauffman, S.A.: The Origins of Order: Self-organization and Selection in Evolution. Oxford University Press, New York (1993)
21. Smith, J.I., Steel, M., Hordijk, W.: Autocatalytic sets in a partitioned biochemical network. J. Syst. Chem. **5**, 2 (2014)
22. Sousa, F.L., Hordijk, W., Steel, M., Martin, W.F.: Autocatalytic sets in E. coli metabolism. J. Syst. Chem. **6**(1), 4 (2015)
23. Hutchison 3rd, C.A., et al.: Design and synthesis of a minimal bacterial genome. Science **351** (aad6280), 6253 (2016)
24. Fraser, C.M., et al.: The minimal gene complement of Mycoplasma genitalium. Science **270** (5235), 397–403 (1995)
25. International Human Genome Sequencing C: Finishing the euchromatic sequence of the human genome. Nature **431**(7011), 931–945 (2004)
26. Kauffman, S.: Metabolic stability and epigenesis in randomly constructed genetic nets. J. Theor. Biol. **22**(3), 437–467 (1969)
27. Murrugarra, D., Laubenbacher, R.: Regulatory patterns in molecular interaction networks. J. Theor. Biol. **288**, 66–72 (2011)
28. Murrugarra, D., Dimitrova, E.S.: Molecular network control through boolean canalization. EURASIP J. Bioinform. Syst. Biol. **2015**(1), 9 (2015)
29. Kauffman, S.: Homeostasis and differentiation in random genetic control networks. Nature **224**(5215), 177–178 (1969)
30. Torres-Sosa, C., Huang, S., Aldana, M.: Criticality is an emergent property of genetic networks that exhibit evolvability. PLoS Comput. Biol. **8**(9), e1002669 (2012)
31. Mukherjee, S.: The Emperor of All Maladies: A Biography of Cancer. Scribner, New York (2011). 1st Scribner trade paperback edn
32. Sole, R.V., Newman, M.: Extinctions and biodiversity in the fossil record. In: Mooney, H.A., Canadell, J.G. (ed) Encyclopedia of Global Environmental Change. The Earth System: Biological and Ecological Dimensions of Global Environmental Change, vol. 2, pp. 297–301. Wiley, Chichester (2002)
33. Kauffman, S.A.: At Home in The Universe: The Search for Laws of Self-organization and Complexity. Oxford University Press, New York (1995)
34. Shmulevich, I., Kauffman, S.A., Aldana, M.: Eukaryotic cells are dynamically ordered or critical but not chaotic. Proc. Natl. Acad. Sci. USA **102**(38), 13439–13444 (2005)
35. Kauffman, S.A.: Requirements for evolvability in complex systems: orderly dynamics and frozen components. Phys. D: Nonlinear Phenom. **42**(1), 135–152 (1990)
36. Wood, R.D.: Human DNA Repair Genes (2014). http://sciencepark.mdanderson.org/labs/wood/dna_repair_genes.html. Accessed 28 Dec 2016
37. Pray, L.: DNA replication and causes of mutation. Nat. Educ. **1**(1), 214 (2008)
38. Rosenfeldt, M.T., Ryan, K.M.: The multiple roles of autophagy in cancer. Carcinogenesis **32** (7), 955–963 (2011)
39. Chin, C.F., Yeong, F.M.: Safeguarding entry into mitosis: the antephase checkpoint. Mol. Cell. Biol. **30**(1), 22–32 (2010)

40. Stracker, T.H., Usui, T., Petrini, J.H.: Taking the time to make important decisions: the checkpoint effector kinases Chk1 and Chk2 and the DNA damage response. DNA Repair (Amst) **8**(9), 1047–1054 (2009)
41. Grivennikov, S.I., Greten, F.R., Karin, M.: Immunity, inflammation, and cancer. Cell **140**(6), 883–899 (2010)
42. Rama, I., Grinyo, J.M.: Malignancy after renal transplantation: the role of immunosuppression. Nat. Rev. Nephrol. **6**(9), 511–519 (2010)
43. Nordling, C.O.: A new theory on cancer-inducing mechanism. Br. J. Cancer **7**(1), 68–72 (1953)
44. Knudson Jr., A.G.: Mutation and cancer: statistical study of retinoblastoma. Proc. Natl. Acad. Sci. USA **68**(4), 820–823 (1971)
45. Limpens, J., et al.: Lymphoma-associated translocation t(14;18) in blood B cells of normal individuals. Blood **85**(9), 2528–2536 (1995)
46. Marculescu, R., Le, T., Simon, P., Jaeger, U., Nadel, B.: V(D)J-mediated translocations in lymphoid neoplasms: a functional assessment of genomic instability by cryptic sites. J. Exp. Med. **195**(1), 85–98 (2002)
47. Pernick, N.L.: How Cancer Arises Based on Complexity Theory (2017). http://www.natpernick.com/HowCancerArises.pdf. Accessed 27 May 2018
48. Lindor, N.M., McMaster, M.L., Lindor, C.J., Greene, M.H., National Cancer Institute DoCPCO, Prevention Trials Research Group: Concise Handbook of Familial Cancer Susceptibility Syndromes, 2nd edn. (2008). J. Natl. Cancer Inst. Monogr. (38), 1–93

Agent Cognition Through Micro-simulations: Adaptive and Tunable Intelligence with NetLogo LevelSpace

Bryan Head[1(✉)] and Uri Wilensky[1,2]

[1] Center for Connected Learning and Computer-Based Modeling, Northwestern Institute of Complex Systems, Department of EECS, Northwestern University, Evanston, IL, USA
bryan.head@u.northwestern.edu, uri@northwestern.edu
[2] Department of Learning Sciences, Northwestern University, Evanston, IL, USA
https://ccl.northwestern.edu/

Abstract. We present a method of endowing agents in an agent-based model (ABM) with sophisticated cognitive capabilities and a naturally tunable level of intelligence. Often, ABMs use random behavior or greedy algorithms for maximizing objectives (such as a predator always chasing after the closest prey). However, random behavior is too simplistic in many circumstances and greedy algorithms, as well as classic AI planning techniques, can be brittle in the context of the unpredictable and emergent situations in which agents may find themselves. Our method, called agent-centric Monte Carlo cognition (ACMCC), centers around using a separate agent-based model to represent the agents' cognition. This model is then used by the agents in the primary model to predict the outcomes of their actions, and thus guide their behavior. To that end, we have implemented our method in the NetLogo agent-based modeling platform, using the recently released LevelSpace extension, which we developed to allow NetLogo models to interact with other NetLogo models. As an illustrative example, we extend the Wolf Sheep Predation model (included with NetLogo) by using ACMCC to guide animal behavior, and analyze the impact on agent performance and model dynamics. We find that ACMCC provides a reliable and understandable method of controlling agent intelligence, and has a large impact on agent performance and model dynamics even at low settings.

Keywords: Agent-based modeling · Artificial intelligence
Agent cognition · Multi-level agent-based modeling · NetLogo

1 Introduction

Agent-based modeling (ABM) has long proven to be a powerful method for simulating complex systems [3,8,15]. Over the last decade, multi-level agent-based modeling (MLABM) has extended this power by enabling researchers to create

A. J. Morales et al. (Eds.): ICCS 2018, SPCOM, pp. 71–81, 2018.
https://doi.org/10.1007/978-3-319-96661-8_7

systems of connected ABMs [9]. This allows one to model a system with multiple components or levels by creating separate ABMs for each component that are then connected. We recently released LevelSpace [5], which brings the ability to integrate many different ABMs to NetLogo [14], one of the most widely used ABM platforms. Here, we demonstrate how to leverage the power of MLABM in order to define sophisticated and tunable cognitive systems for guiding agent behavior.

Agent cognition differs from classic artificial intelligence algorithms in that the goal is not to generate the most optimal course of action. Instead, it is desirable to have a level of intelligence that is appropriate for the agents in question. For instance, agents representing humans should have significantly different cognitive capabilities than agents representing sheep, which should, in turn, have significantly different capabilities than agents representing ants. Furthermore, the task is complicated by the fact that, due to the subject matter of most ABMs, agents exist in complex environments with many other agents, such that the patterns of their world are an emergent result of the collective actions of those agents. Making matters more difficult, agents are typically only aware of local information. Thus, agents will find themselves in surprising and unexpected circumstances, and must continue to act reasonably in those circumstances. Hence, our goal is to design a method of agent cognition in which:

1. Agents are given goals and information about their local state, and they determine what actions to take.
2. Agents have a tunable level of "intelligence".
3. What agents know about the world and how they think the world works is definable. In other words, agents should often have simplified understandings of the world.
4. Agents behave reasonably when in unexpected circumstances, or when something surprising occurs.
5. Ideally, the agents' cognitive processes should be inspectable and understandable. Researchers should be able to understand why agents are doing what they are doing.

Our method, agent-centric Monte Carlo cognition (ACMCC), accomplishes these goals by defining an ABM that represents how an agent thinks the world works. By leveraging MLABM, agents then use this cognitive ABM to make short term predictions about what will happen when they choose different actions. The agents will then make their final action decision based on the predicted outcomes of each action.

2 Related Work

Under one perspective, an ABM defines a Markov chain, a small part of which is observable by each agent. Hence, the framework of partially observable markov decision processes (POMDPs) may be applicable in this context [7]. Indeed, MDPs have been used to model decision making of agents in dynamic environments in other works. For instance, [1] combines POMDPs with reinforcement learning to perform dynamic decision making for a single agent in a stochastic environment. Further, [6] extends MDPs to work with multiple agents via Markov games. MDPs are often solved by computing the sum of discounted future reward values at each state (via dynamic programming). However, such strategies are infeasible in the context of ABMs as, typically, even the local area which an agent can observe has a vast state-space, due to the number of agents and the fact that agents are often defined by continuous attributes, and the rate at which new information is encountered. Furthermore, even if it were possible to find the optimal solution to the POMDPs involved, it would not be desirable to do so, as agents typically have bounded rationality by design. That said, our work may be characterized as using Monte Carlo simulations to approximate the solutions to POMDPs of each agents' surroundings.

Monte Carlo tree search (MCTS) has proven to be a particularly powerful method to address decision processes, especially in the context of games [2]. Notably, MCTS was combined with deep reinforcement learning techniques to create the first super-human Go playing program [11,12]. Fundamentally, MCTS works by using Monte Carlo "playouts" to determine the expected value of each possible move the player can take. It then uses those expected values to guide future playouts, so that they focus on more promising paths. The technique may be further augmented with other methods (such as deep neural networks estimating the values of states) to further guide the playouts. In this work, we take a similar Monte Carlo approach to sampling the eventual value of actions, but use a pure Monte Carlo strategy instead. Section 7 discusses extending this work to incorporate advancements in MCTS.

[4] also demonstrated a method of defining agent cognition via MLABM using an early version of LevelSpace. In that work, agents sent information about their surroundings to simple neural network models. The output of the neural networks would then determine what action the agents would take. This work has several advantages compared to that, though they are complimentary methods, as discussed in Sect. 7. First, this method requires no training. Second, this method uses explicit cognitive representations. That is, the objects of the agents' cognition are of the same form as the objects of the world (entities in an ABM). When a sheep is considering the actions of a wolf, a wolf agent exists in their cognitive model. Furthermore, we can directly observe why an agent does what it does, and what it expects to happen for different actions. This is in sharp contrast with a neural network, in which the agent's knowledge is implicitly embedded in the weights of the neurons. Finally, this method offers a parameter that directly corresponds with the cognitive capabilities of the agents: namely, the number of simulations the agent runs. In contrast, the number and size of the layers in a neural network controls the cognitive capabilities of the agents, but how that translates into concrete differences in cognition and behavior is opaque.

3 Method

ACMCC works as follows:

1. The modeler defines what actions an agent can take (e.g., turn left, turn right, go forward, eat, etc.), what the agent knows/believes (e.g., what the agent can see), what the agent is trying to do (e.g., maximize food intake while staying alive), and how the agent thinks the world works via a cognitive model defined by a separate agent-based model. That is, this cognitive model may be a simplified version of the main ABM, which operates according to mechanisms as understood by the agent. Thus, the agent should be able to use this cognitive model to make predictions about what will happen in the full model.
2. During each tick of the simulation, each agent runs a settable number of rollouts, or short simulations of their surroundings, using its cognitive model, with initial conditions based on their surroundings. During these rollouts, the agent selects random actions and tracks how well they meet their objectives as a consequence of those actions. The agent's performance is evaluated based on reward: during a rollout, the agent will receive a reward based on what happens (the modeler defines the reward based on the agent's objectives).
3. The agent then selects an action based on the results of these rollouts.

See Fig. 2 for an example of these rollouts.

A significant advantage of this method is that it gives researchers several tunable parameters that precisely control agents' "intelligence", such as the number of rollouts to run and the length of each rollout. Having such control over agents' intelligence allows modelers to, for instance, naturally adjust agents' cognitive capabilities based on what is reasonable for those agents, or have an evolvable "intelligence" parameter that directly correlates to the agents' cognitive capabilities. Note that we do not claim that this is how the actual cognitive processes of the modeled agents work. Rather, this method simply gives modelers a practical method of controlling the sophistication of agents' decision making abilities.

4 Model

In order to demonstrate our method, we extend the Wolf Sheep Predation model found in the NetLogo Models Library [13]. Wolf Sheep Predation is a basic predator-prey model consisting of wolves, sheep, and grass, run on a two dimensional toroidal world. Wolves and sheep are point-like agents that move in continuous space, while grass exists as a binary value on a grid (either grass is present at a particular cell or it is not). In the original model, wolves eat sheep, sheep eat grass, and grass regrows at a fixed rate. When agents eat, they acquire energy, and when agents move, they lose a small amount of energy. Wolves and sheep reproduce based on a probability each tick, splitting their energy with their child. All movement is random: each tick, agents turn left or right by a random angle less than 45° and take a step forward of unit length. If an agent

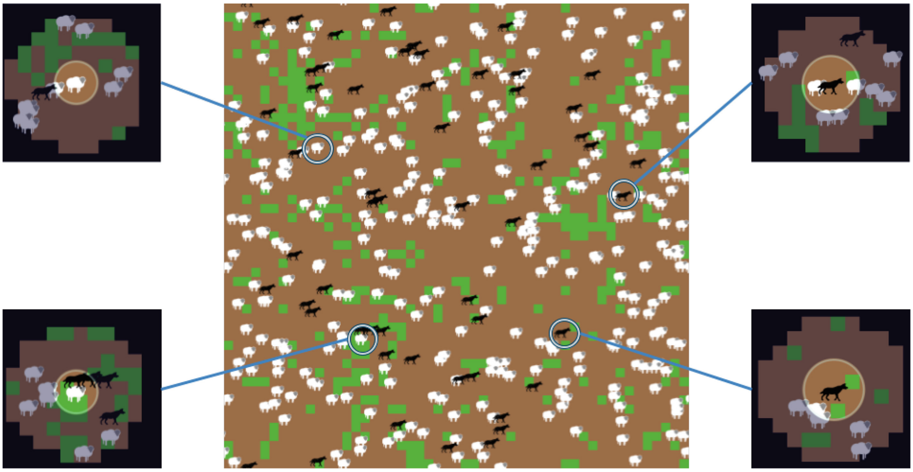


Fig. 1. A state from the modified Wolf Sheep Predation model. The initial state of the cognitive model for each of the highlighted agents is shown.

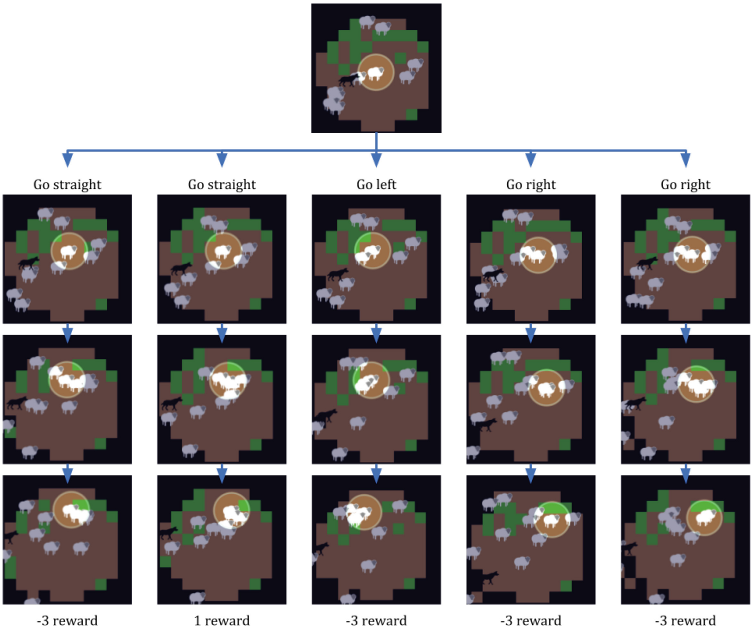


Fig. 2. The tree of rollouts performed by the highlighted sheep, which is facing north. The rollouts progress from top to bottom, with the first row showing the state after 1 tick, and the last row showing the final state. The sheep's initial actions and final reward values are shown (without discounting).

runs out of energy, it dies and is removed from the world. If an agent is eaten, it dies and is removed from the world.

In our extension of the model, the random movement is replaced with ACMCC. Agents have the following actions available to them:

1. Turn right 30°, take a step forward of unit length, and eat if possible.
2. Take a step forward of unit length, and eat if possible.
3. Turn left 30°, take a step forward of unit length, and eat if possible.

Reproduction still occurs based on a probability.

In order to decide what actions to take each tick, agents use a simple cognitive model. First, agents are given a vision radius; their cognitive model will be initialized with what they see in this radius each tick. The cognitive model is a simplified version of the original Wolf Sheep Predation. The world represented is much smaller than the full model: a little bit larger than their vision radius. It is initialized with the agents surroundings: the positions and headings of the surroundings wolves and sheep, as well as the positions of the live and dead grass, and finally the position, heading, and energy of the agent itself. Furthermore, this model only includes factors of which the agents would plausibly be aware of. The cognitive model thus makes the following simplifications:

- Only the energy of the primary agent (called the ego) is tracked. This is because the ego cannot observe the energy of the other agents.
- There is no reproduction. However, our model could be further extended to make reproduction an action, and have the child's well-being incorporated into the rollout reward-function. It was left out for simplicity and because that it is reasonable to think that the possibility of suddenly reproducing is not factored into the decision of which piece of grass to go for.
- There is no grass regrowth. Again, the ego cannot observe the regrowth state of the grass, and these are short-term simulations.
- There is nothing outside of the ego's vision radius. Should an agent move past the vision radius during a rollout, the grass at that point is filled in randomly based on the observed density of grass.

All agents in the cognitive model act randomly. While there are many interesting possibilities for selecting actions of agents in rollouts, such as further embedding cognitive models (thus giving agents a kind of theory of mind), or refining rollout decision making based on past rollouts (as in MCTS) we chose to begin with the simplest approach: random action in rollouts. Reward in the rollouts is calculated as the agents change in energy, with a discounting factor, so that energy acquired in the first tick counts more than energy acquired in the second, and so forth. Furthermore, death is given an automatic reward value of -1000.

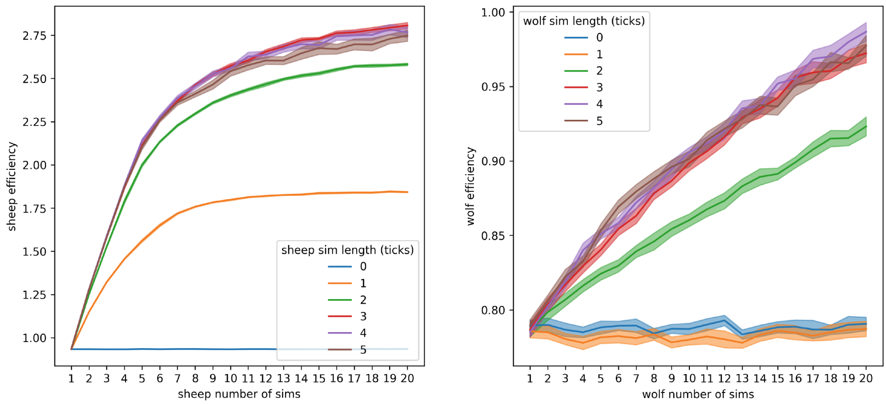
The main predator-prey model and cognitive model are then combined as follows. Each tick of the predator-prey model proceeds as in the original Wolf Sheep Predation, save that agents send information about their surroundings to their cognitive model, which then runs a set number of rollouts of a set length, and reports the resulting reward values. The agent then takes the action that had the highest mean reward. See Fig. 1 for a state of the predator-prey model and the corresponding initial state of several of the cognitive models. Figure 2 shows the various rollouts performed by the cognitive model of a sheep at that tick. Through these rollouts, the sheep finds that the two sheep coming from the east almost always eat the patch of grass to the northeast, and the sheep coming from the north almost always eat the grass to the north-northwest. Thus, the only grass the sheep manages to successfully eat is the northernmost (final image in the second column). Thus, the sheep decides to go straight. In this way, ACMCC allows for agents to perform sophisticated decision making where a greedy algorithm would have failed (because the closest grass would have been eaten by other sheep first).

5 Results

In order to understand the effects of the cognitive model on Wolf Sheep Predation, several experiments were performed. In each experiment, one agent-type had the number and length of its rollouts varied while the other agent-type was kept random. The efficiency of the agents in various tasks as well as population levels were recorded. Two measurements of efficiency were used: one for sheep and one for wolves. Sheep efficiency was measured as the ratio of the amount of grass eaten in a tick to the amount of grass we would expect to be eaten by random sheep, given grass density and number of sheep: `grass-eaten / (num-sheep * grass-density)`. Similarly, wolf efficiency was measured as the ratio of the amount of sheep eaten in a tick to the amount of sheep we would expect to be eaten by random wolves, given sheep density and number of wolves: `sheep-eaten / (num-wolves * sheep-density)`. This measure was used as it controls for population dynamics: it measures how well the agents are doing compared to how well we would expect random agents to be doing. Thus, in both cases, an efficiency of 1 is random. However, as there are relatively few wolves (in the base model, wolves oscillate between 50 and 100, while sheep oscillate between 120 and 200) with relatively few prey (the sheep's prey, grass, covers around 1000 locations at any given tick), this number rests below 1 for the baseline case (at around 0.76) due to the fact that wolves and their prey are discrete entities. Sheep, on the other hand, are more numerous with an evenly spread food source, and thus their baseline rests close to the theoretical value of 1, at 0.93.

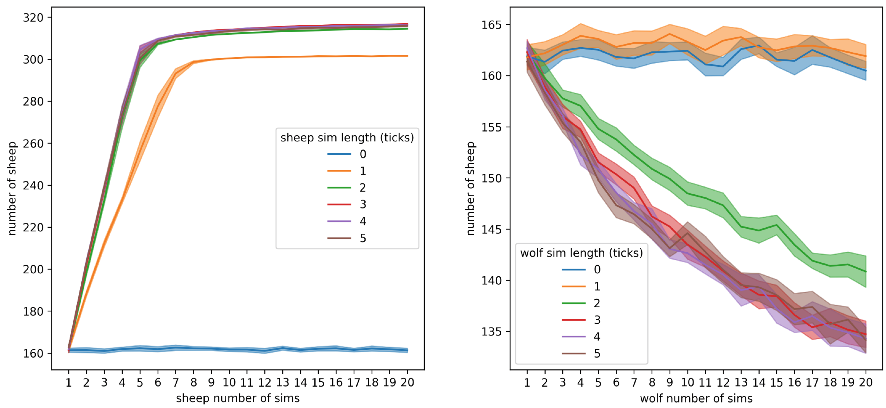
For both wolves and sheep, number of rollouts was varied from 1 to 20 (note that performing only a single rollout is equivalent to random behavior) and the length of rollouts was varied from 0 to 5. Vision radius was set to 5 for both agent types. Default parameter settings (as found in the NetLogo Models Library) were used for all other parameters. Runs were carried out for 2,000 ticks. 20 repetitions were performed for each combination of parameters. Figure 3 shows the efficiency results for both agent-types.

Finally, in order to begin to understand the impact of the cognitive model on the dynamics of the model, we examine the sheep population. Using the same



(a) Sheep using rollouts, wolves random (b) Wolves using rollouts, sheep random

Fig. 3. Mean efficiency of sheep and wolves at eating their respective foods (grass and sheep) for various combinations of number and length of rollouts. Bands show 95% confidence intervals. Note that the y-axes have different scales.



(a) Sheep using rollouts, wolves random (b) Wolves using rollouts, sheep random

Fig. 4. Mean sheep population for various combinations of number and length of rollouts performed by sheep and wolves. Bands show 95% confidence intervals. Note that the y-axes have different scales.

runs as above, Fig. 4(a) shows the mean population of sheep for each combination of number and length of sheep rollouts. In contrast, Fig. 4(b) shows mean sheep population in response to wolf cognitive abilities. Means are taken from tick 500 onwards in each run, to allow for the system to reach stability.

6 Discussion

We find that the number of simulations has a direct and monotonic impact on agent performance, and thus works well as an "intelligence" parameter. Furthermore, even a small number of short rollouts dramatically increases performance in sheep. Thus, it is possible to give hundreds of agents fairly sophisticated short-term reasoning in this way without too high of a performance impact: all simulations performed here are runnable in real-time on a standard, modern laptop. Rollout length appeared to reach maximum efficacy at around three ticks. This is not surprising; the longer rollouts are run for, the less accurate they will be, especially as agents could only see to distance 5. This invites the question of how to do longer term predictions about the future; this is discussed in Sect. 7.

While wolves were not helped as much as sheep, their performance was still improved with a relatively small number of short rollouts. Regardless, their poorer performance is not surprising: catching a moving prey is much more difficult than catching a stationary one, or even than avoiding being caught. Nevertheless, wolf performance is impressive considering the branching factor and their poor model of sheep behavior. Methods for overcoming their natural disadvantage are discussed in Sect. 7.

As shown in Fig. 4b, as the wolves become smarter, their food supply significantly drops. This is particularly significant as fewer sheep means fewer sheep reproducing, and thus less food for the wolves being introduced each tick. This is unlike the sheep's food source, which regrows at each location at a fixed rate, and thus, the fewer locations occupied by grass, the faster it is introduced into the system. The effect of wolf behavior on sheep population highlights the difficulties of applying cognitive systems; due to the aggregate-level feedback loop in the system, what's beneficial to the individual can be harmful to the group.

More broadly, these results indicate that this is a promising strategy for giving agents more sophisticated, yet tunable cognitive capabilities in a natural way, using nothing but agent-based modeling.

7 Future Work

While this work lays a solid foundation for a novel approach to agent cognition, it can be extended in many interesting ways.

First, as the agents are only performing a handful of rollouts, it is important that those rollouts focus on promising/likely actions and futures. This is particularly salient in the case of the wolves, who have the more difficult task of catching a moving target. There are a number of established ways of accomplishing this. First, in MCTS, this is accomplished by using the results of past rollouts to guide future rollouts. There are some difficulties here: with the few number of rollouts and the continuous state-space of many ABMs, it is unlikely for the rollouts to encounter the same state twice. Regardless, this method is immediately useful for the choice of the first action, and could be modified to work on future actions by either compressing states (using a learned encoder) or scoring sequences of actions, regardless of state.

Another method that has recently proven to be highly effective is combining MCTS with neural networks, as was done in AlphaGo [11] and AlphaGo Zero [12]. In ABM, neural networks could be trained to both select likely actions for the ego, as well as to predict the behavior of the other agents. This method would both improve the efficacy of rollouts, and offer a way of incorporating learning into the system, if that is desirable. Our initial work in this direction has been promising, which combines this work with the work in [4].

Another method of better predicting the actions of other agents would be to embed another layer of ACMCC models inside the first layer. That is, agents would have a kind of theory of mind, where they try to emulate the thinking of the other agents based on what they can observe. The drawback to the naive implementation of this, however, is that it is performance intensive (initial experiments reinforce this). However, this may be circumvented by doing a kind of MCTS for each agent at the high level; that is, each agent in the cognitive model uses the results of past rollouts to improve their behavior in future rollouts, thus emulating the cognition of the other agents without performing any additional rollouts. Regardless, the naive approach of further embedding cognitive models could be quite effective for models with fewer agents that require highly sophisticated reasoning. For instance, [10] examines an approach to simulating theory of mind in a multi-agent simulation somewhat along these lines.

Another challenge is adapting this method to work with continuous action spaces. As described here, the actions must be discrete; wolves and sheep can only turn by fixed amounts, while they can turn by a continuous amount in the original model. This could be accomplished by interpolating between different discrete parameters for an action.

While the method as applied here appears to be effective for short-term reasoning, it does not perform any kind of long term reasoning. A simple way of adapting the method to perform longer term reasoning would be to decrease the accuracy of the simulation while increasing it's speed by changing its timestep. Another method would be to have the cognitive model operate on a coarser grain than the main model.

Thus, while this work lays the foundation for sophisticated agent cognition, it opens up many possible avenues of exploration as well.

References

1. Barve, S.S.: Dynamic decision model using partially observable Markov decision process. Int. J. Emerg. Trend Eng. Basic Sci. **2**(1), 785–788 (2015)
2. Browne, C.B., et al.: A survey of monte carlo tree search methods. IEEE Trans. Comput. Intell. AI Games **4**(1), 1–43 (2012)
3. Epstein, J.M.: Generative Social Science: Studies in Agent-Based Computational Modeling. Princeton University Press, Princeton (2006)
4. Head, B., Hjorth, A., Brady, C., Wilensky, U.: Evolving agent cognition with Netlogo LevelSpace. In: Proceedings of the Winter Simulation Conference (2015)
5. Hjorth, A., Head, B., Wilensky, U.: LevelSpace NetLogo extension. Center for Connected Learning and Computer Based Modeling, Northwestern University, Evanston, IL (2015). http://ccl.northwestern.edu/levelspace/index.html

6. Littman, M.L.: Markov games as a framework for multi-agent reinforcement learning. In: Cohen, W.W., Hirsh, H. (eds.): Machine Learning Proceedings 1994, pp. 157–163. Morgan Kaufmann, San Francisco (CA) (1994)
7. Lovejoy, W.S.: A survey of algorithmic methods for partially observed Markov decision processes. Ann. Oper. Res. **28**(1), 47–65 (1991)
8. Macal, C.M., North, M.J.: Agent-based modeling and simulation: ABMS examples. In: Winter Simulation Conference, WSC 2008, pp. 101–112. IEEE (2008)
9. Morvan, G.: Multi-level agent-based modeling: A Literature Survey. CoRR abs/1205.0 (2013)
10. Rabinowitz, N.C., Perbet, F., Song, H.F., Zhang, C., Eslami, S.M.A., Botvinick, M.: Machine Theory of Mind. arXiv:1802.07740 [cs], February 2018
11. Silver, D., et al.: Mastering the game of Go with deep neural networks and tree search. Nature **529**(7587), 484–489 (2016)
12. Silver, D., et al.: Mastering the game of Go without human knowledge. Nature **550**(7676), 354–359 (2017)
13. Wilensky, U.: NetLogo Wolf Sheep Predation Model. Center for Connected Learning and Computer-Based Modeling, Northwestern University, Evanston, IL (1997). http://ccl.northwestern.edu/netlogo/models/WolfSheepPredation
14. Wilensky, U.: NetLogo. Center for Connected Learning and Computer-Based Modeling, Northwestern University, Evanston, IL (1999). http://ccl.northwestern.edu/netlogo/
15. Wilensky, U., Rand, W.: An Introduction to Agent-Based Modeling: Modeling Natural, Social, and Engineered Complex Systems with NetLogo. MIT Press, Cambridge (2015)

Check for updates

Recognizing Complex Behavior Emerging from Chaos in Cellular Automata

Gabriela M. González[1(✉)], Genaro J. Martínez[1,2], M. A. Aziz Alaoui[3], and Fangyue Chen[4]

[1] Artificial Life Robotics Lab, Escuela Superior de Cómputo, Instituto Politécnico Nacional, Mexico City, Mexico
gmorenog1400@alumno.ipn.mx
[2] Unconventional Computing Lab, University of the West of England, Bristol BS16 1QY, UK
genaro.martinez@uwe.ac.uk
[3] Normandie Univ, UNIHAVRE, LMAH, FR-CNRS-3335, ISCN, BP. 540, 76600 Le Havre, France
aziz.alaoui@univ-lehavre.fr
[4] School of Science, Hangzhou Dianzi University, Hangzhou, China
fychen@hdu.edu.cn

Abstract. In this research, we explain and show how a chaotic system displays non-trivial behavior as a complex system. This result is reached modifying the chaotic system using a memory function, which leads to a new system with elements of the original function which are not evident in a first step. We proof that this phenomenology can be apprehended selecting a typical chaotic function in the domain of elementary cellular automata to discover complex dynamics. By numerical simulations, we demonstrate how a number of gliders emerge in this automaton and how some controlled subsystems can be designed within this complex system.

Keywords: Complex dynamics · Chaos · Emergence · Gliders
Glider guns · Memory

1 Preliminaries

As far as we know a classification scheme of complex systems is missing, although, in the case of elementary cellular automata, several approaches have been described in recent decades. Nevertheless, this is a difficult problem that is initially determined as an undecidable problem in the context of the theory of cellular automata [9]. However, several approaches have been considered (we can refer for example to [1,3,12,17,31] and references cited therein). These researches show that, to date, all approaches using elementary cellular automata, do not match, for more details, see [14].

In this paper, we study a classic chaotic elementary cellular automaton and, using a memory function, we describe elements of non-trivial behaviour which

A. J. Morales et al. (Eds.): ICCS 2018, SPCOM, pp. 82–90, 2018.
https://doi.org/10.1007/978-3-319-96661-8_8

emerge as patterns. The last are named gliders, particles, waves, or mobile self-localizations. This non-trivial behaviour is qualified as *complex*. Specialized and historic books recommended to introduce to complex systems theory are [4–6, 19, 21]. For cellular automata theory, we refer to [1, 18, 27, 29].

Here, we study a particular case with elementary cellular automaton rule 126, classified by Wolfram as a chaotic rule in [29, 30]. Reviewing and composing the original function with a memory function we will discover non-trivial patterns. The original study about this analysis was published in [15]. This way, the present paper is an extension on its controllability and codification of gliders in the evolution rule with memory and how scale complex its behaviour by glider collisions. The paper has the next structure. Section two introduces basic concepts. Section three displays a description of rule 126 and its chaotic behaviour. Section four explains how works a memory function in cellular automata. Section five shows the non-trivial behaviour in rule 126 with memory.

1.1 Basic Notation

One-dimensional cellular automata is represented by an array of *cells* x_i where $i \in Z$ and each x takes a value from a finite alphabet Σ. Thus, a sequence of cells $\{x_i\}$ of finite length n describes a string or *global configuration* c on Σ. The set of finite configurations will be expressed as Σ^n. An evolution is comprised by a sequence of configurations $\{c_i\}$ produced by the mapping $\Phi : \Sigma^n \rightarrow \Sigma^n$; thus the global relation is symbolized as: $\Phi(c^t) \rightarrow c^{t+1}$, where t represents time and every global state of c is defined by a sequence of cell states. The global relation is determined over the cell states in configuration c^t updated at the next configuration c^{t+1} simultaneously by a local function φ as follows: $\varphi(x^t_{i-r}, \ldots, x^t_i, \ldots, x^t_{i+r}) \rightarrow x^{t+1}_i$.

Wolfram represents one-dimensional cellular automata with two parameters (k, r), where $k = |\Sigma|$ is the number of states, and r is the neighbourhood radius. This way, elementary cellular automata domain is defined by parameters $(2, 1)$. There are Σ^n different neighbourhoods (where $n = 2r + 1$) and k^{k^n} distinct evolution rules. The evolutions in this paper have periodic boundary conditions.

Conventional cellular automata are ahistoric (memoryless): i.e., the new state of a cell depends on the neighbourhood configuration solely at the preceding time step of φ. Thus, cellular automata with *memory* can be considered as an extension of the standard framework of cellular automata where every cell x_i is allowed to remember some period of its previous evolution. Basically memory is based on the state and history of the system, thus we design a memory function ϕ, as follows: $\phi(x^{t-\tau}_i, \ldots, x^{t-1}_i, x^t_i) \rightarrow s_i$, such that $\tau < t$ determines the backwards degree of memory and each cell $s_i \in \Sigma$ is a function of the series of states in cell x_i up to time-step $t - \tau$. Finally to execute the evolution we apply the original rule again as follows: $\varphi(\ldots, s^t_{i-1}, s^t_i, s^t_{i+1}, \ldots) \rightarrow x^{t+1}_i$.

In cellular automata with memory, while the mapping φ remains unaltered, a historic memory of past iterations is retained by featuring each cell as a summary of its previous states; therefore cells *canalize* memory to the map φ. As an example, we can take the memory function ϕ as a *majority memory*: $\phi_{maj} \rightarrow s_i$, where in case of a tie given by $\Sigma_1 = \Sigma_0$ in ϕ, we shall take the last value x_i. So ϕ_{maj} represents the classic majority function for three variables [20] on cells $(x_i^{t-\tau}, \ldots, x_i^{t-1}, x_i^t)$ and defines a temporal ring before calculating the next global configuration c. In case of a tie, it is allowed to break it in favor of zero if $x_{\tau-1} = 0$, or to one whether $x_{\tau-1} = 1$.

The representation of a elementary cellular automata with memory is given as follows: $\phi_{CARm:\tau}$, where CAR represents the decimal notation of a particular elementary cellular automata rule and m the kind of memory given with a specific value of τ. Thus, the majority memory (maj) working in elementary cellular automaton rule 126 checking tree cells ($\tau = 3$) of history is simply denoted as $\phi_{R126maj:3}$.

Note that memory is as simple as any cellular automata, and that the global behaviour produced by the local rule is totally unpredictable, it can lead to emergent properties and so be complex. Memory functions were developed and extensively studied by Sanz in [24–26]. Memory in elementary cellular automata have been studied, showing its potentiality to report complex behaviour from chaotic systems and beyond in [11,12,16], and recently in [7] authors have included hybrid versions. Thus, we can conjecture that a memory function can report complex behaviour as follows: $f_{chaos}(\phi) \rightarrow f_{complex}$.

2 Elementary Cellular Automaton Rule 126

The local-state transition function φ corresponding to rule 126 displays a high concentration of states 1s. This way, $\varphi_{R126} = \{1 \text{ if } 110, 101, 100, 011, 010, 001; 0 \text{ if } 111, 000\}$.

Rule 126 has a chaotic global behaviour typical from class III in Wolfram's classification [29]. In φ_{R126} we can easily recognize an initial high probability of alive cells, i.e. cells in state '1'; with a 75% to appear in the next time and, complement of only 25% to get a state 0.

Figure 1 shows these cases in typical snapshots of rule 126. Evolving from a single cell in state '1' yield a patters like a Sierpinski triangle (Fig. 1a). From a 50% random initial configuration, we can see an unordered evolution

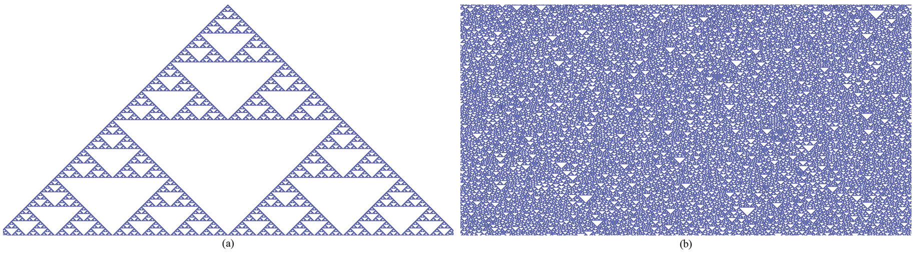


Fig. 1. Dynamics in elementary cellular automaton rule 126. (a) Sierpinski triangle is the evolution in its global dynamics from a single cell in state 1. (b) Second snapshot is calculated from a random initial density to 50%. Both space-time diagrams evolve on a ring of 1,000 cells for 512 generations.

without observing any recognizable pattern (Fig. 1b). Both evolutions induce that evolution rule 126 display a chaotic global behaviour. To explore carefully a lot of these properties we will analyze its basin of attractors in the next section.

2.1 Basins of Attraction

A basin (of attraction) field of a finite cellular automata is the set of basins of attraction into which all possible states and trajectories will be organized by the local function φ. The topology of a single basin of attraction may be represented by a diagram, the *state transition graph*. Thus, the set of graphs composing the field specifies the global behaviour of the system [27].

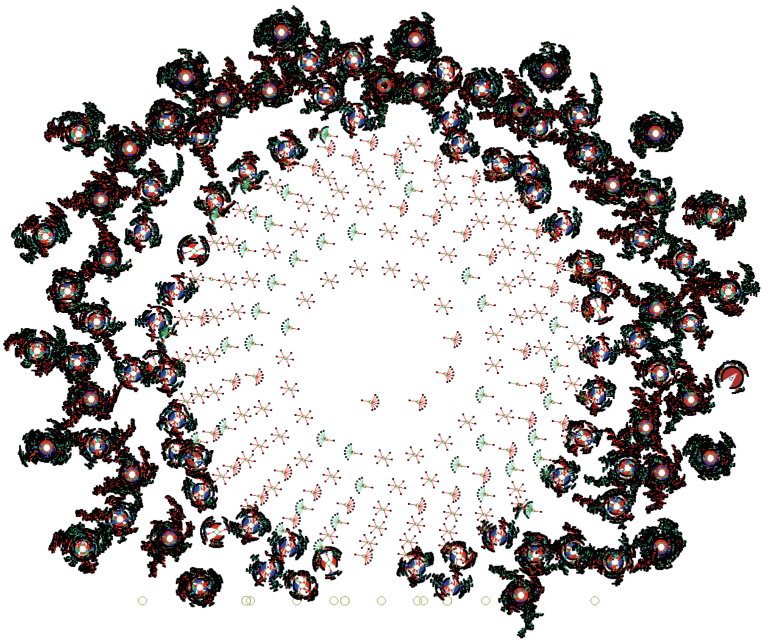


Fig. 2. Basin of attractors for rule 126 and a length of 20 cells.

Generally, a basin can also recognizes cellular automata with chaotic or complex behaviour following previous results on attractors which has been reported by Wuensche and Lesser in [27] for rings of length 2 to 15. Thus, Wuensche characterizes the Wolfram's classes as a basin classification for chaos and complexity in [27].

The basin depicted in Fig. 2 shows the whole set of non-equivalent basins in rule 126 for ring equal to 20 cells.[1] Particularly in this basin we can see that attractors have not long transients or long periodic attractors, but several of them have low in-degree and low leaf density. A quick observation is

[1] Basins and attractors were calculated with *Discrete Dynamical System* DDLab available from http://www.ddlab.org/.

that these basin of attractors are symmetric in their leafs, that induce chaos [8]. Moreover, at the same time we can see some non symmetric attractors and some of them have moderate transients that could induce a non-trivial complex behaviour inside the chaos. A study discussing particularly the complex behaviour by graphs in elementary cellular automata is available in [13].

3 Dynamics Emerging with Memory

This section presents the results of selecting a majority memory (maj) with $\tau = 4$ in rule 126 deriving a new function names $\phi_{R126maj:4}$ [15]. Figure 3 displays an evolution for the rule $\phi_{R126maj:4}$, showing its complex behaviour.

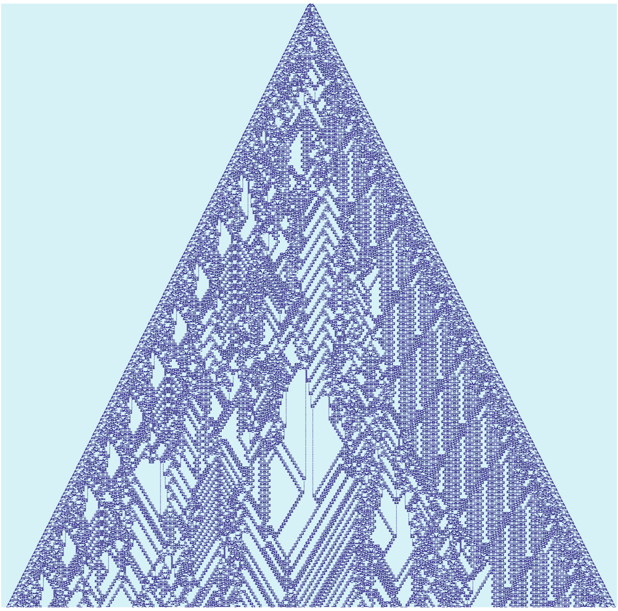


Fig. 3. Complex behaviour emerging in rule 126 with memory, rule $\phi_{R126maj:4}$. Initial configuration 111001 in a ring of 1,000 cells to 1,000 times. The evolution is filtered to get a better view of gliders and collisions. The universe of non-trivial patterns emerge including stationary particles, gliders and several glider guns. Interesting collisions from this initial condition include solitons, annihilations, reflexions, fusions and more. The initial condition was selected intentionally to produce a non-symmetric evolution.

4 Collisions of Gliders

In this section, we have done a systematic analysis of multiple collisions of gliders in $\phi_{R126maj:4}$. The next table presents an equation for every collision and its result. Figure 4 illustrates explicitly these simulations.

Equation	Result	Equation	Result	Equation	Result
$e_1(p_1) + g_1(p_1) + e_1(p_1) + g_2(p_1) + e_1(p_1)$	0	$e_1(p_1) + g_1(p_1) + e_1(p_1) + s_1(p_1) + e_1(p_1)$	s1	$e_1(p_2) + g_1(p_3)^2 + e_1(p_2) + g_2(p_4) + e_1(p_1)$	g_1
$e_1(p_1) + g_1(p_1) + e_1(p_1) + g_2(p_2) + e_1(p_2)$	g_1	$e_1(p_1) + g_1(p_1) + e_1(p_1) + s_1(p_2) + e_1(p_2)$	s_1	$e_1(p_2) + g_1(p_3)^2 + e_1(p_2) + g_2(p_5) + e_1(p_2)$	g_1
$e_1(p_1) + g_1(p_1) + e_1(p_1) + g_2(p_3) + e_1(p_2)$	$g_3 + g_3$	$e_1(p_2) + g_1(p_2) + e_1(p_2) + s_1(p_1) + e_1(p_1)$	s_1	$e_1(p_1) + g_1(p_4)^2 + e_1(p_1) + g_2(p_1) + e_1(p_1)$	g_1
$e_1(p_1) + g_1(p_1) + e_1(p_1) + g_2(p_4) + e_1(p_1)$	$g_4 + g_4$	$e_1(p_2) + g_1(p_2) + e_1(p_2) + s_1(p_2) + e_1(p_2)$	s_1	$e_1(p_1) + g_1(p_4)^2 + e_1(p_1) + g_2(p_2) + e_1(p_2)$	g_1^2
$e_1(p_1) + g_1(p_1) + e_1(p_1) + g_2(p_5) + e_1(p_2)$	g_2	$e_1(p_2) + g_1(p_3) + e_1(p_1) + s_1(p_1) + e_1(p_1)$	s_1	$e_1(p_1) + g_1(p_4)^2 + e_1(p_1) + g_2(p_3) + e_1(p_2)$	$g_3 + g_3$
$e_1(p_2) + g_1(p_2) + e_1(p_2) + g_2(p_1) + e_1(p_1)$	g_2	$e_1(p_2) + g_1(p_3) + e_1(p_1) + s_1(p_2) + e_1(p_2)$	s_1	$e_1(p_1) + g_1(p_4)^2 + e_1(p_1) + g_2(p_4) + e_1(p_1)$	g_2
$e_1(p_2) + g_1(p_2) + e_1(p_2) + g_2(p_2) + e_1(p_2)$	0	$e_1(p_1) + g_1(p_4) + e_1(p_2) + s_1(p_1) + e_1(p_1)$	s_1	$e_1(p_1) + g_1(p_4)^2 + e_1(p_1) + g_2(p_5) + e_1(p_2)$	g_1
$e_1(p_2) + g_1(p_2) + e_1(p_2) + g_2(p_3) + e_1(p_2)$	g_1	$e_1(p_1) + g_1(p_4) + e_1(p_2) + s_1(p_2) + e_1(p_2)$	s_1	$e_1(p_2) + g_1(p_5)^2 + e_1(p_2) + g_2(p_1) + e_1(p_1)$	g_1
$e_1(p_2) + g_1(p_2) + e_1(p_2) + g_2(p_4) + e_1(p_1)$	$g_3 + g_3$	$e_1(p_2) + g_1(p_5) + e_1(p_2) + s_1(p_1) + e_1(p_1)$	s_1	$e_1(p_2) + g_1(p_5)^2 + e_1(p_2) + g_2(p_2) + e_1(p_2)$	g_1
$e_1(p_2) + g_1(p_2) + e_1(p_2) + g_2(p_5) + e_1(p_2)$	$g_4 + g_4$	$e_1(p_2) + g_1(p_5) + e_1(p_2) + s_1(p_2) + e_1(p_2)$	s_1	$e_1(p_2) + g_1(p_5)^2 + e_1(p_2) + g_2(p_3) + e_1(p_2)$	g_1^2
$e_1(p_2) + g1_(p_3) + e_1(p_1) + g_2(p_1) + e_1(p_1)$	$g_4 + g_4$	$e_2(p_5) + g_3(p_2) + e_1(p_2) + g_4(p_2) + e_2(p_5)$	s_2	$e_1(p_2) + g_1(p_5)^2 + e_1(p_2) + g_2(p_4) + e_1(p_1)$	$g_3 + g_3$
$e_1(p_2) + g_1(p_3) + e_1(p_1) + g_2(p_2) + e_1(p_2)$	g_2	$e_1(p_1) + g_3(p_4) + e_2(p_2) + s_2(p_2) + e_2(p_2)$	g_4	$e_1(p_2) + g_1(p_5)^2 + e_1(p_2) + g_2(p_5) + e_1(p_2)$	g_2
$e_1(p_2) + g_1(p_3) + e_1(p_1) + g_2(p_3) + e_1(p_2)$	0	$e_1(p_1) + g_1(p_1)^2 + e_1(p_1) + g_2(p_1) + e_1(p_1)$	g_2	$e_1(p_1) + g_1(p_1) + e_1(p_1) + s_1(p_1) + e_1(p_1) + g_2(p_1) + e_1(p_1)$	s_1
$e_1(p_2) + g_1(p_3) + e_1(p_1) + g_2(p_4) + e_1(p_1)$	g_1	$e_1(p_1) + g_1(p_1)^2 + e_1(p_1) + g_2(p_2) + e_1(p_2)$	g_1	$e_1(p_1) + g_1(p_1) + e_1(p_1) + s_1(p_1) + e_1(p_1) + g_2(p_2) + e_1(p_2)$	s_1
$e_1(p_2) + g_1(p_3) + e_1(p_1) + g_2(p_5) + e_1(p_2)$	$g_3 + g_3$	$e_1(p_1) + g_1(p_1)^2 + e_1(p_1) + g_2(p_3) + e_1(p_2)$	g_1	$e_1(p_1) + g_1(p_1) + e_1(p_1) + s_1(p_1) + e_1(p_1) + g_2(p_3) + e_1(p_2)$	s_1
$e_1(p_1) + g_1(p_4) + e_1(p_1) + g_2(p_1) + e_1(p_1)$	$g_3 + g_3$	$e_1(p_1) + g_1(p_1)^2 + e_1(p_1) + g_2(p_4) + e_1(p_1)$	g_1^2	$e_1(p_1) + g_1(p_1) + e_1(p_1) + s_1(p_1) + e_1(p_1) + g_2(p_4) + e_1(p_1)$	s_1
$e_1(p_1) + g_1(p_4) + e_1(p_1) + g_2(p_2) + e_1(p_1)$	$g_4 + g_4$	$e_1(p_1) + g_1(p_1)^2 + e_1(p_1) + g_2(p_5) + e_1(p_2)$	$g_3 + g_3$	$e_1(p_1) + g_1(p_1) + e_1(p_1) + s_1(p_1) + e_1(p_1) + g_2(p_5) + e_1(p_2)$	s_1
$e_1(p_1) + g_1(p_4) + e_1(p_1) + g_2(p_3) + e_1(p_1)$	g_2	$e_1(p_2) + g_1(p_2)^2 + e_1(p_2) + g_2(p_1) + e_1(p_1)$	$g_3 + g_3$	$e_1(p_1) + g_1(p_1) + e_1(p_1) + s_1(p_2) + e_1(p_2) + g_2(p_1) + e_1(p_1)$	s_1
$e_1(p_1) + g_1(p_4) + e_1(p_1) + g_2(p_4) + e_1(p_1)$	0	$e_1(p_2) + g_1(p_2)^2 + e_1(p2) + g_2(p2) + e_1(p2)$	g_2	$e_1(p_1) + g_1(p_1) + e_1(p_1) + s_1(p_2) + e_1(p_2) + g_2(p_2) + e_1(p_2)$	s_1
$e_1(p_1) + g_1(p_4) + e_1(p_1) + g_2(p_5) + e_1(p_1)$	g_1	$e_1(p_2) + g_1(p_2)^2 + e_1(p_2) + g_2(p_3) + e_1(p_2)$	g_1	$e_1(p_1) + g_1(p_1) + e_1(p_1) + s_1(p_2) + e_1(p_2) + g_2(p_3) + e_1(p_2)$	s_1
$e_1(p_2) + g_1(p_5) + e_1(p_1) + g_2(p_2) + e_1(p_2)$	$g_3 + g_3$	$e_1(p_2) + g_1(p_2)^2 + e_1(p_2) + g_2(p_5) + e_1(p_2)$	g_1^2	$e_1(p_1) + g_1(p_1) + e_1(p_1) + s_1(p_2) + e_1(p_2) + g_2(p_5) + e_1(p_2)$	s_1
$e_1(p_2) + g_1(p_5) + e_1(p_1) + g_2(p_3) + e_1(p_2)$	$g_4 + g_4$	$e_1(p_2) + g_1(p_3)^2 + e_1(p_2) + g_2(p_1) + e_1(p_1)$	g_1^2	$e_1(p_1) + g_1(p_1) + e_1(p_1) + g_2(p_1) + g_4(p_4) + e_2(p_2)$	g_4
$e_1(p_2) + g_1(p_5) + e_1(p_1) + g_2(p_4) + e_1(p_1)$	g_2	$e_1(p_2) + g_1(p_3)^2 + e_1(p_2) + g_2(p_2) + e_1(p_2)$	$g_3 + g_3$	$e_1(p_1) + g_1(p_1) + e_1(p_1) + g_2(p_2) + g_4(p_5) + e_2(p_3)$	g_4
$e_1(p_2) + g_1(p_5) + e_1(p_1) + g_2(p_5) + e_1(p_2)$	0	$e_1(p_2) + g_1(p_3)^2 + e_1(p_2) + g_2(p_3) + e_1(p_2)$	g_2	$e_1(p_1) + g_1(p_1) + e_1(p_1) + g_2(p_3) + g_4(p_1) + e_2(p_4)$	$g_4 + g_3 + g_3$

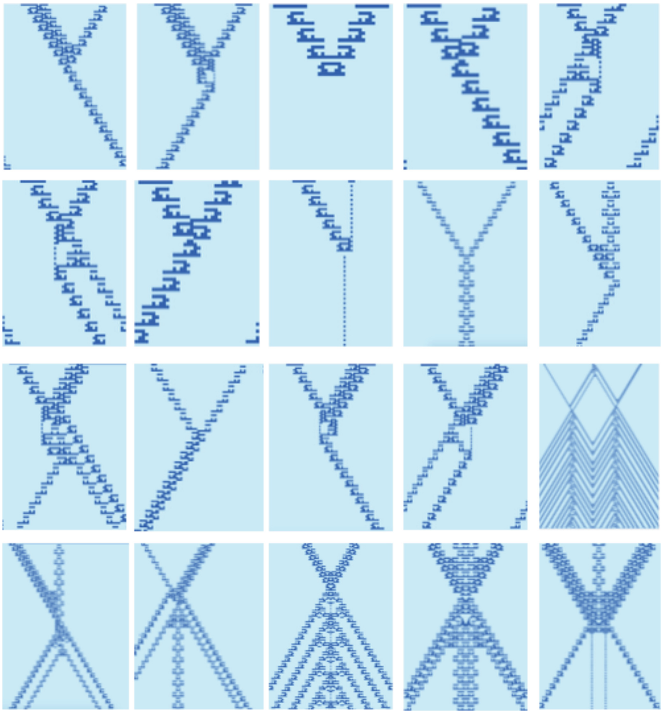


Fig. 4. Codifying gliders in $\phi_{R126m:\tau}$ to controlling collisions. Some examples are illustrated in this figure where we can see annihilations, self-organization, glider production, and glider guns.

5 Conclusions

Memory is a useful tool to discover complexity in dynamical systems from composed functions. In this paper, we have extended these results previously to controller gliders in rule 126 systematically. The next step of our research will design more complex constructions including computable devices by glider collisions.

References

1. Adamatzky, A.: Identification of Cellular Automata. Taylor and Francis, London (1994)
2. Adamatzky, A. (ed.): Collision-Based Computing. Springer, London (2002)
3. Adamatzky, A., Martínez, G.J.: On generative morphological diversity of elementary cellular automata. Kybernetes **39**(1), 72–82 (2010)
4. Aziz-Alaoui, M., Bertelle, C.: From System Complexity to Emergent Properties. Springer, Heidelberg (2009)
5. Bar-Yam, Y.: Dynamics of Complex Systems. Addison-Wesley, New York (1997)

6. Boccara, N.: Modeling Complex Systems. Springer, New York (2003)
7. Chen, B., Chen, F., Martínez, G.J.: Glider collisions in hybrid cellular automaton with memory rule (43,74). Int. J. Bifurc. Chaos **27**(6) (2017). https://doi.org/10.1142/S0218127417500821
8. Chen, G., Dong, X.: From Chaos to Order. In: World Scientific Series on Nonlinear Science, Series A, vol. 24 (1998)
9. Culik II, K., Yu, S.: Undecidability of CA classification schemes. Complex Syst. **2**(2), 177–190 (1988)
10. Kauffman, S.A.: The Origins of Order: Self-organization and Selection in Evolution. Oxford University Press, New York (1993)
11. Martínez, G.J., Adamatzky, A., Sanz, R.A., Mora, J.C.S.T.: Complex dynamic emerging in rule 30 with majority memory a new approach. Complex Syst. **18**(3), 345–365 (2010)
12. Martínez, G.J., Adamatzky, A., Sanz, R.A.: Designing complex dynamics with memory. Int. J. Bifurc. Chaos **23**(10), 1330035–131 (2013)
13. Martínez, G.J., Adamatzky, A., Chen, B., Chen, F., Mora, J.C.S.T.: Simple networks on complex cellular automata: from de Bruijn diagrams to jump-graphs. In: Zelinka, I., Chen, G. (eds.) Swarm Dynamics as a Complex Network, pp. 241–264. Springer, Heidelberg (2018)
14. Martínez, G.J.: A note on elementary cellular automata classification. J. Cell. Autom. **8**(3–4), 233–259 (2013)
15. Martínez, G.J., Adamatzky, A., Mora, J.C.S.T., Sanz, R.A.: How to make dull cellular automata complex by adding memory: rule 126 case study. Complexity **15**(6), 34–49 (2012)
16. Martínez, G.J., Adamatzky, A., Sanz, R.A.: Complex dynamics of cellular automata emerging in chaotic rules. J. Nonlinear Syst. Appl. **22**(2) (2012). https://doi.org/10.1142/S021812741250023X
17. McIntosh, H.V.: Wolfram's class IV and a good life. Physica D **45**, 105–121 (1990)
18. McIntosh, H.V.: One Dimensional Cellular Automata. Luniver Press, Bristol (2009)
19. Mainzer, K., Chua, L.: The Universe as Automaton: From Simplicity and Symmetry to Complexity. Springer, Heidelberg (2012)
20. Minsky, M.: Computation: Finite and Infinite Machines. Prentice Hall, Englewood Cliffs (1967)
21. Mitchell, M.: Complexity: A Guided Tour. Oxford University Press, New York (2009)
22. Martínez, G.J., McIntosh, H.V., Mora, J.C.S.T., Vergara, S.V.C.: Determining a regular language by glider-based structures called phases f_i_1 in rule 110. J. Cell. Autom. **3**(3), 231–270 (2008)
23. Prokopenko, M., Michael Harré, M., Lizier, J.T., Boschetti, F., Peppas, P., Kauffman, S.: Self-referential basis of undecidable dynamics: from The Liar Paradox and The Halting Problem to The Edge of Chaos, CoRR abs/1711.02456 (2017)
24. Sanz, R.A., Martin, M.: Elementary CA with memory. Complex Syst. **14**, 99–126 (2003)
25. Sanz, R.A.: Elementary rules with elementary memory rules: the case of linear rules. J. Cell. Autom. **1**, 71–87 (2006)
26. Sanz, R.A.: Cellular Automata with Memory. Old City Publishing, Philadelphia (2009)
27. Wuensche, A., Lesser, M.: The Global Dynamics of Cellular Automata. Addison-Wesley Publishing Company, Reading (1992)
28. Wolfram, S.: Universality and complexity in cellular automata. Physica D **10**, 1–35 (1984)

29. Wolfram, S.: Cellular Automata and Complexity. Addison-Wesley Publishing Company, Englewood Cliffs (1994)
30. Wolfram, S.: A New Kind of Science. Wolfram Media Inc., Champaign (2002)
31. Wuensche, A.: Classifying cellular automata automatically. Complexity **4**(3), 47–66 (1999)

Simulation of Scale-Free Correlation in Swarms of UAVs

Shweta Singh(✉) and Mieczyslaw M. Kokar

Department of Electrical and Computer Engineering,
Northeastern University, Boston, MA 02115, USA
{ssingh,mkokar}@ece.neu.edu

Abstract. Natural phenomena such as flocking in birds, known as *emergence*, is proved to be scale-invariant, i.e., flocks of birds exhibit scale-free correlations which give them the ability to achieve an effective collective response to external conditions and environment changes to survive under predator attacks. However, the role of scale-free correlations is not clearly understood in artificially simulated systems and thus more investigation is justifiable. In this paper, we present an attempt to mimic the scale-free behavior in swarms of autonomous agents, specifically in Unmanned Aerial Vehicles (UAVs). We simulate an agent-based model, with each UAV treated as a dynamical system, performing persistent surveillance of a search area. The evaluation results show that the correlation in swarms of UAVs can be scale-free. Since this is a part of ongoing research, open questions and future directions are also discussed.

Keywords: Scale-free correlation · Collective behavior · Emergence UAVs · Dynamical systems

1 Introduction

Self-organized collective behaviors are observed in a variety of natural and artificial systems. Collective behavior is a way of showing how a group as a whole reacts to changes in its environment. Flocking of birds is one such example [7]. For instance, birds protect themselves from a predator attack by making random movements in a flock which confuses the attacker. This self-organized property of the system, like flocking, is also known as *emergent property*. Emergent properties appear when a number of simple components (agents) in a system operate and interact in an environment, forming more complex dynamic behaviors as collective [1]. These emerge from integration between the components of the system via their local interactions.

Cavagna et al. in their paper [2] suggests that collective response can be achieved by scale-free correlations. The authors calculated the behavioral (velocity and heading) correlations in a group of starling flocks and proved that this correlation is scale-free, i.e., the range of spatial correlation does not have a constant value but it scales linearly with the size of the flock. This result is important

A. J. Morales et al. (Eds.): ICCS 2018, SPCOM, pp. 91–97, 2018.
https://doi.org/10.1007/978-3-319-96661-8_9

as this provides a group of birds to effectively respond to any perturbation in the environment. Scale-free correlation gives a long and effective perception range to individual agent which is much larger than the direct inter-individual (agent-agent) interaction range [4]. This means that any small change in one agent causes change in the whole group. Cavagna et al. studied the correlated domains in real starling flocks. But, the role of scale-free correlation is still unclear in artificial systems [5]. One such attempt is performed by Viscek et al. [10] with a very simple model of particles moving with a constant absolute velocity. The authors investigated the emergence of flocking as the self-organized behavior in systems of particles with a single interaction rule. The paper does not calculates scale-free correlation but showed that the average absolute velocity of the group scales with some factor of perturbation in the system. In this paper, we extend the flocking model to swarms of Unmanned Aerial Vehicles (UAVs) by adding multiple rules and proves that such flocks of UAVs also follows scale-free correlation as the real starling flocks in [2].

With the advent in technology, smaller UAVs are replacing the bigger ones as swarms of collaborating autonomous UAVs, performing a specific mission, can solve problems better than collections of UAVs that are controlled centrally. However, studies have also shown that swarms of UAVs can exhibit unpredictable emergent behaviors due to interactions or external environmental influence. To design a mechanism to learn or control such behaviors, much can be learned from nature as natural systems, such as starling flocks are capable of achieving an effective collective behavior. The scale-free correlations give robustness and adaptivity to the system. But, there is still a gap between the capabilities of these robotic UAV swarms and those of natural swarms, especially when it concerns performing collectively intelligent tasks. In this paper, we describe simulations of emergent behaviors in swarms of UAVs for a specific mission of persistent surveillance of a search area in two-dimensional environment and study whether such swarms exhibit scale-free correlations as in nature. Our aim is to leverage the scale-free correlation in swarms of UAVs to avoid undesirable emergent behaviors. This paper presents just part of the work we are doing to study emergent behaviors in swarms of UAVs. Section 2 discusses the agent-based model for UAVs followed by results and analysis of scale-free correlation in Sect. 3. We discuss the results and future work at the end of this paper in Sect. 4.

2 Agent-Based Model for UAVs

In our simulations, each UAV, U, operates in a two dimensional environment, $E \subset X \times Y$, where $X = Y = \mathbb{N}$ are natural numbers that enumerate the *cells* (coordinate locations) of the environment. We consider a plume, $P \subseteq E$, as the targeted search area to which U provides persistent surveillance by measuring the environment at each time instant $t \in T = \mathbb{N}$. Each measurement covers a subset of the environment $S(t) \subset E$, termed as *sensor footprint*. Assuming mission starts at $t_s \in T$, initial time, and ends at $t_f \in T$, the final time, a measurement is a sequence $S(t_s), S(t_s + 1), \ldots, S(t_f)$. Such sequences of measurements are considered as *paths* that each UAV (sensor) travels. The path is controlled by selecting acceleration, a, at each time from an interval, $[a_{min}, a_{max}]$, which we refer to as the control input for the UAV. The behavioral state for each UAV, U_i, is given by its location and velocity vector, i.e., $[\mathbf{x_i}, \mathbf{v_i}]$. The quality of the surveillance provided by UAVs is measured by a metric called *information age*, I_{age}, of the plume (see [8] for more details). For every time instant, I_{age} is updated by following the equation: $I_{age}(P, t) = \sum_{i,j} A_{ij}(t), (i, j) \in P$, where A_{ij} is an age value at (i, j) cell. The goal of each UAV is to provide persistent surveillance of the plume by minimizing the information age metric. The optimization problem for this scenario is formalized in [8].

To simulate flocking like behaviors in the swarms of UAVs, we implemented three behavioral rules of Reynold's Boids [7], i.e., *Cohesion* (each UAV moves in the direction of the centroid of the neighboring UAVs), *Separation* (maintain minimum distance with neighboring UAV (neighborhood is defined by sensor footprint)), and *Alignment* (heading of each UAV is equal to the average heading of the neighboring UAVs). These rules are shown graphically in Fig. 1.

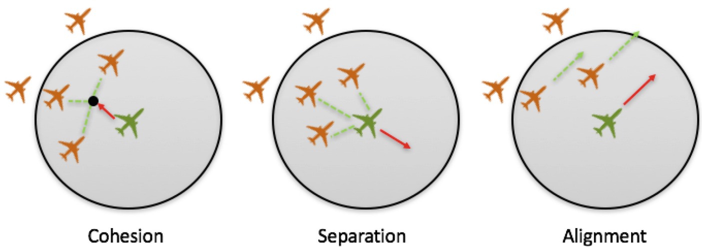


Fig. 1. Three behavioral rules for UAVs Interactions. The gray circles represent the sensor footprints of the UAVs at the center (green plane). All the other UAVs inside the sensor footprint define the neighborhood of the center UAV (shown as the green dashed line). The red arrow represents the possible next move for the green UAV based on the selected behavioral rule.

3 Results

3.1 Simulation of UAVs Flock

For simulations of UAV agents, we used NetLogo [9] as the platform for experimentation as it allows modeling of an agent like entities in a simulation environment supported by a scripting language, visual animator, and data output mechanisms. Figure 2 shows a snapshot of NetLogo simulation environment. Here, with the specific parameter values (sensor-radius = 10, minimum-separation-threshold = 1, max-align-turn = 35, max-cohere-turn = 6.0, max-separate-turn = 10.9, mean-speed = 1.0), UAVs are showing flocking behavior while preserving the information age metric (Sect. 2). In other words, not only swarms of UAVs exhibit the emergent flocking behavior due to the inclusion of Boids rules [6],

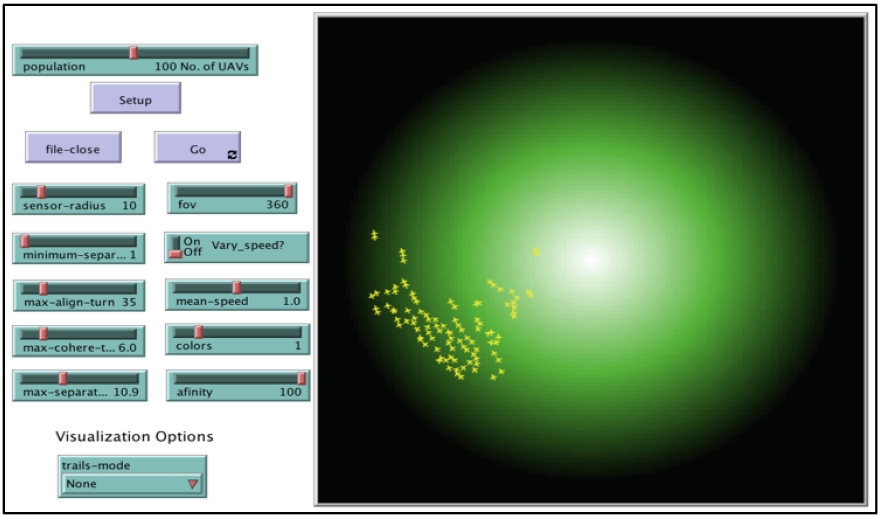


Fig. 2. NetLogo simulation environment. The environment, 120 × 120 grid, containing the circular plume (green cells) of radius 30 units. The yellow agents (in shape of planes) show individual UAVs flying in the two-dimensional environment forming a flock. The total number of UAV agents in the current simulation is set to 100 with. NetLogo allows a run-time change in these parameter values using slider bars shown on the left.

but also this flock remains inside the plume to minimize the overall information age in order to solve the optimization problem [8]. In this scenario of persistent surveillance, this emergent behavior is undesirable as the information age can be reduced faster with separated movements of UAVs rather than as a single flock. We are not showing the plot of the information age in this paper, refer to [8] for detailed analysis. Our aim here is to study correlation in behavioral states of UAVs in a flock. Next subsection discusses this analysis.

3.2 Analysis of Scale-Free Correlation

Following the approach described in [2], we consider the correlation metric that measures the fluctuations in the velocities of UAVs in a UAV flock. The fluctuation vector, $\mathbf{u}_i$, of an individual UAV$_i$ is calculated by subtracting the average velocity vector from the velocity vector of the UAV$_i$. The velocities are calculated by averaging over several instance of time in the flocking event.

$$\mathbf{u}_i = \mathbf{v}_i - \frac{1}{N} \sum_{k=1}^{N} \mathbf{v}_k \tag{1}$$

where N is the number of UAVs in a flock. The correlation function [2], $Corr(r)$ is given as:

$$Corr(r) = \frac{\sum_{ij} \mathbf{u_i} \cdot \mathbf{u}_j \delta(r - r_{ij})}{\sum_{ij} \delta(r - r_{ij})} \tag{2}$$

where r_{ij} is mutual distance between UAV_i and UAV_j. The delta function is defined as:

$$\delta(r - r_{ij}) = \begin{cases} 1 & r = r_{ij} \\ 0 & \text{otherwise} \end{cases} \tag{3}$$

Figure 3 shows that the correlation decays with increasing r and eventually becomes negative at larger distances. The distance for which this correlation function is zero is called *correlation length* (ξ), i.e.,

$$C(r = \xi) = 0 \tag{4}$$

Cavagna et al. [2] empirically showed that this correlation length is proportional to flock size, L, (the maximum distance between any two agents in a flock), i.e.,

$$\xi(aL) = a\xi(L) \tag{5}$$

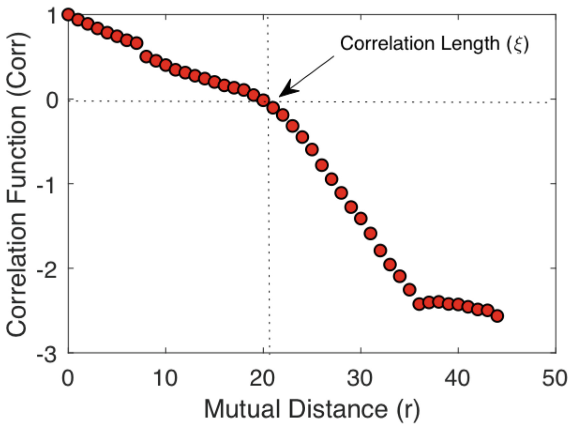


Fig. 3. Correlation Function (Corr) versus Mutual Distance (r). The correlation function showed here determines the fluctuations of UAV's velocities. We calculated the average value over multiple instances of time. The plot showed that it decays and crosses zero and becomes negative. The distance at which correlation is zero is correlation length (ξ), here this value is around 20. The simulation is for N = 100 UAVs.

In other words, the correlation length grows with the flock's size, L, with a proportionality constant, a. This relationship suggests that the correlation of velocity fluctuations becomes a scale-free (power-law) structure. If the correlation size never depends on the flock size, then there is no characteristic scale for

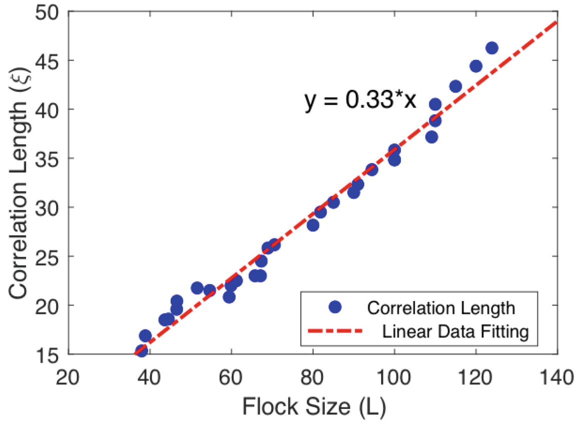


Fig. 4. Correlation Length (ξ) versus Flock Size (L). Each point represents flocking event in simulations with varying sizes of UAV. The constant of proportionality obtained is 0.33, which is almost the same as the empirical value.

the correlation size. This is an intuitive explanation for why above equation is related to the power law. For a real starling flock, [2] found the value of a to be 0.35. Some other papers also try to find this value for simulated flocks. One such example is [5] where they found this value to be around 0.38.

For UAV simulations, we plotted the correlation length versus varying sizes of UAV flock. This relationship is shown in Fig. 4. We ran multiple simulations for $N = 30$, 50, 100 and 150 UAVs. The plot shows that the correlation length and the flock size are clearly correlated. After performing the linear regression, we found the value of the gradient to be around 0.33 for a UAV flock (shown as the red line in Fig. 4). It is interesting to see that the value of the constant a obtained experimentally is almost the same as the value found in the empirical work of Cavagna et al. [2]. The goodness of fit is evaluated using R-square and adjusted R-square statistics [3] which came out to be 0.9792 and 0.9785, respectively, for the current fit. In other words, this linear fit explains about 97% of the total variations in the data about the average. The value of the adjusted R-square closer to 1 indicates a better fit.

4 Discussion and Future Work

Cavagna et al. [2] state that when correlation length is larger than the interaction range in self-organized groups, the scale-free correlations causes qualitatively different collective behaviors, i.e., flocking in starling flocks. We also saw this in the UAV persistent surveillance scenario. Flocking emerges out as an emergent behavior which is undesirable for the current scenario as flocking causes poor surveillance of the plume.

From the short experiment presented in this paper, we found that the correlation of velocities in our artificially simulated scenario of UAVs follows scale-free power law as in the real starling flocks. Correlation length at which the correlation value is zero indicates the existence of at least more than one correlated domains of UAVs, and also points out to the position in the flock where it is uncorrelated. We proved, experimentally, that this correlation length scales with flock's size, i.e., the larger the flock, the larger the correlation length. Learning about these facts gives us the ability to understand the structure of the UAV flock, and allows us to exploit this information to control the movement of the flock. So the question here is: Can this scale-free correlation help us to separate correlated domains so that we can avoid this type of undesirable behaviors? This is the question that we wanted to raise in this current work and pursue in next extensions.

Some of the other high-level questions that we are targeting in our research of emergent behaviors: What is the role of scale in emergent behaviors in UAV flocks? Under what conditions UAVs exhibit scale-free correlations? Do such emergent behaviors are always desirable or can they have an adversarial effect on the performance of the system? Though we have not answered all the targeted questions in the current work, but it does open up the possibility of learning from such scale-free correlated behaviors and apply them to control the undesirable emergent behaviors in swarms of autonomous agents such as UAVs.

References

1. Bonabeau, E., Dessalles, J.-L.: Detection and emergence. arXiv preprint arXiv:1108.4279 (2011)
2. Cavagna, A., Cimarelli, A., Giardina, I., Parisi, G., Santagati, R., Stefanini, F., Viale, M.: Scale-free correlations in starling flocks. Proc. Natl. Acad. Sci. **107**(26), 11865–11870 (2010)
3. Glantz, S.A., Slinker, B.K.: Primer of applied regression and analysis of variance. Number Sirsi) i9780070234079 (1990)
4. Khaluf, Y., Ferrante, E., Simoens, P., Huepe, C.: Scale invariance in natural and artificial collective systems: a review. J. R. Soc. Interface **14**(136) (2017). https://doi.org/10.1098/rsif.2017.0662
5. Niizato, T., Murakami, H., Gunji, Y.-P.: Emergence of the scale-invariant proportion in a flock from the metric-topological interaction. Biosystems **119**, 62–68 (2014)
6. Reynolds, C.: Boids. https://www.red3d.com/cwr/boids/. Accessed 7 Feb 2018
7. Reynolds, C.W.: Flocks, herds and schools: a distributed behavioral model. ACM SIGGRAPH Comput. Graph. **21**(4), 25–34 (1987)
8. Singh, S., Lu, S., Kokar, M.M., Kogut, P.A.: Detection and classification of emergent behaviors using multi-agent simulation framework (WIP). In: Proceedings of the Symposium on Modeling and Simulation of Complexity in Intelligent, Adaptive and Autonomous Systems, p. 3. Society for Computer Simulation International (2017)
9. Tisue, S., Wilensky, U.: NetLogo: design and implementation of a multi-agent modeling environment. In: Proceedings of Agent, pp. 7–9 (2004)
10. Vicsek, T., Czirók, A., Ben-Jacob, E., Cohen, I., Shochet, O.: Novel type of phase transition in a system of self-driven particles. Phys. Rev. Lett. **75**(6), 1226 (1995)

Complexity of Maxmin-ω Cellular Automata

Ebrahim L. Patel(✉)

Mathematical Institute, University of Oxford, Oxford OX2 6GG, UK
ebrahim.patel@maths.ox.ac.uk

Abstract. We present an analysis of an additive cellular automaton (CA) under asynchronous dynamics. The asynchronous scheme employed is maxmin-ω, a deterministic system, introduced in previous work with a binary alphabet. Extending this work, we study the impact of a varying alphabet size, i.e., more than the binary states often employed. Far from being a simple positive correlation between complexity and alphabet size, we show that there is an optimal region of ω and alphabet size where complexity of CA is maximal. Thus, despite employing a fixed additive CA rule, the complexity of this CA can be controlled by ω and alphabet size. The flavour of maxmin-ω is, therefore, best captured by a CA with a large number of states.

1 Introduction

The maxmin-ω system was introduced in [9] as a model of asynchronous dynamics on networks. Each node in this system updates its state upon receiving a proportion ω of inputs from neighbourhood nodes. Cellular automata are reknowned for their modelling capabilities of a variety of complex systems – be they biological, computational or physical. CA consist of a lattice of identical automata, or "cells", where each cell takes one of a finite set of states. Classical CA update their state synchronously according to some local rule (see [13] for example); asynchrony adds more realism to these models [1,11]. In terms of a CA application, the main attraction for employing maxmin-ω is that it is asynchronous yet deterministic. Moreover, these local interactions of maxmin-ω provide a simpler and intuitive mechanism for asynchrony, mimicking the dynamics of similar models whose applications include neuronal networks [7] and virus transmission [12]. Taking these points together, maxmin-ω looks like a new member of the class of *threshold models* that have their roots in epidemic spreading [2].

Consider a one-dimensional CA lattice, where the neighbourhood $\mathcal{N}_i$ of cell i of radius r is the set $\{i-r, \ldots, i-1, i, i+1, \ldots, i+r\}$, as introduced in [13]. Maxmin-ω views the CA lattice as a network, whose nodes play the role of cells. Thus, a cell state is updated at the end of a *cycle*. The processes that constitute such a cycle are as follows. First, the neighbourhood nodes $\mathcal{N}_i$ complete their k^{th} cycle and then transmit their CA state to node i; the transmission of such a state from node j to i takes *transmission time* $\tau_{ij}(k)$. Node i waits for a

A. J. Morales et al. (Eds.): ICCS 2018, SPCOM, pp. 98–107, 2018.
https://doi.org/10.1007/978-3-319-96661-8_10

fraction ω of the arriving states before processing its new CA state, which takes *processing time* $\xi_i(k+1)$. Once this is complete, the node updates its CA state and simultaneously transmits this state to downstream nodes, where the cycles are reiterated.

We denote the time of state change of node i by $x_i(k+1)$, whilst the CA state of node i in the same cycle is denoted $s_i(k+1)$. Thus, the $(k+1)^{\text{th}}$ update time of node i is given by the following recurrence relation.

$$x_i(k+1) = x_{(\omega)}(k) + \xi_i(k+1) \tag{1}$$

where $x_{(\omega)}(k)$ represents the k^{th} time of arrival of the ω^{th} input from the neighbourhood of i, which we define as the last of the fraction ω of inputs arriving at i; if k is clear from context, we denote this $x_{(\omega)}$ for short. If there are n nodes in the neighbourhood of i, then $x_{(\omega)}$ practically represents the time of arrival of the m^{th} input where $m = \lceil \omega n \rceil$.

For our study, we employ an additive CA rule. We first consider an *alphabet* of CA states taking size Z, namely $\Sigma = \{0, 1, 2, \ldots, Z-1\}$. We represent the (CA) state of the system at cycle $k \in \mathbb{N}$ by the vector $\boldsymbol{s}(k) = (s_1(k), s_2(k), \ldots, s_N(k))$. Suppose a cell is contained in a neighbourhood of size $2r+1$; then a CA rule is a function $f : \{0,1\}^{2r+1} \rightarrow \{0,1\}$ given by $s_i(k+1) = f(\mathcal{N}(s_i(k)))$, where $\mathcal{N}(s_i(k))$ denotes the CA states of $\mathcal{N}_i$ in cycle k. Further, consider

$$\mathcal{A}_i(k) = \{j \in \mathcal{N}_i : x_j(k) + \tau_{ij} \leq x_{(\omega)}(k)\} \tag{2}$$

which is the set of all nodes whose CA states arrive before or at the same time as the ω^{th} input at node i. We call $\mathcal{A}_i(k)$ the set of *affecting nodes of* i. Thus, we focus on the following CA rule

$$s_i(k+1) = \sum_{j \in \mathcal{A}_i} s_j(k) \mod Z. \tag{3}$$

Simply put, the CA state of each cell will be the sum of the fastest arriving states at that cell.

We exhibited the impact of maxmin-ω on CA with a binary alphabet in [9], i.e., $\Sigma = \{0, 1\}$. In this paper, we demonstrate the difference in effect for an extended alphabet, i.e., $|\Sigma| > 2$. In particular, we ask the question: what is the effect of maxmin-ω on the complexity of cellular automata?

2 Cellular Automata with General Alphabet

2.1 Cellular Automaton Pattern Complexity

To classify our cellular automata in space-time, we use the entropy measures of Marr and Hütt in [6]. The Shannon entropy S relies on the density $p(s_j)$ of the CA state $s_j \in \Sigma$ in the time series of the evolving CA states of each cell. Thus, the Shannon entropy of cell i is defined as

$$S_i = -\sum_{j=1}^{|\Sigma|} p(s_j) \log_2 p(s_j). \tag{4}$$

The quantity we require is the Shannon entropy of the overall CA space-time pattern, defined as the average of S_i over the N cells in the lattice: $S = (1/N)\sum_{i=1}^{N} S_i$.

The word entropy W depends on the occurrence of blocks of constant states of length l (l-words) in the time series of a cell, independent of the state comprising them. Thus, if $p(l)$ is the density of an l-word along the time series of a cell i, then

$$W_i = -\sum_{l=1}^{T} p(l) \log_2 p(l) \tag{5}$$

where T is the length of the time series. The word entropy of the entire CA pattern is then defined as the average of W_i over the N cells: $W = (1/N)\sum_{i=1}^{N} W_i$.

For the additive CA rule (3) that we consider, each state is equally likely throughout the evolution of the CA [3]. Then we have that $p(s_j) = 1/Z$, giving

$$S_i = -\sum_{j=1}^{Z} \frac{1}{Z} \log Z = \log Z. \tag{6}$$

Taking the average of this quantity over N cells gives $S = \log Z$. The Shannon entropy is therefore expected to increase logarithmically with the size of alphabet.

As for the word entropy W, it is reliant on the state in a time series of a cell being unchanged over some fixed length l of time. Let $s_j^{(t)}$ denote the state of cell j at time $t \in \mathbb{R}$. Thus, an l-word satisfies the following.

$$s_j^{(t)} = s_j^{(t+1)} = \cdots = s_j^{(t+l-1)} \tag{7}$$

where $s_j^{(t-1)} \neq s_j^{(t)}$ and $s_j^{(t)} \neq s_j^{(t+l)}$. Since all CA states are equally likely, $p\left(s_j^{(t)}\right) = 1/Z$ for all t. The probability of the next state being the same is $P\left(s_j^{(t)} = s_j^{(t+1)}\right) = 1/Z$, whilst the probability of the next state being different is $P\left(s_j^{(t)} \neq s_j^{(t+1)}\right) = (Z-1)/Z$. Then the probability $p(l)$ of observing an l-word is $P\left(s_j^{(t-1)} \neq s_j^{(t)} = s_j^{(t+1)} = \cdots = s_j^{(t+l-1)} \neq s_j^{(t+l)}\right)$, calculated as

$$\frac{Z-1}{Z} \times \underbrace{\frac{1}{Z} \times \cdots \times \frac{1}{Z}}_{l \text{ times}} \times \frac{Z-1}{Z} = \frac{(Z-1)^2}{Z^{l+2}}. \tag{8}$$

The inset of Fig. 1 plots $p(l)$ as a function of Z and l; 1-words in particular are expected to be most frequent. Substituting $p(l)$ into (5) gives W_i; taking the average over all cells gives $W = W_i$ here. Figure 1 plots W as a function of $|\Sigma|$. The word entropy is, thus, expected to decrease as the alphabet size increases.

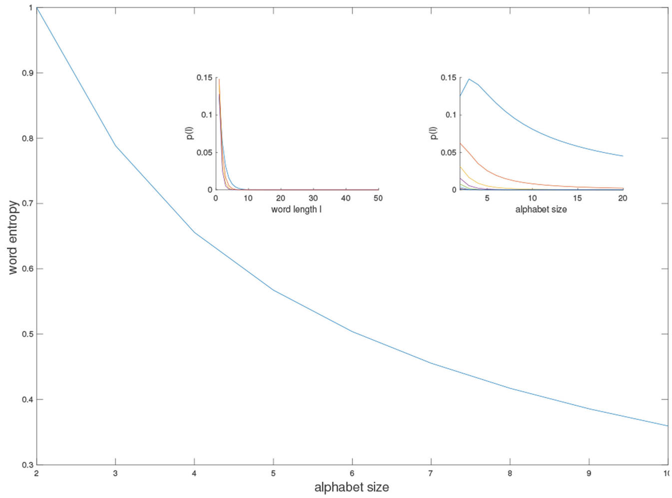


Fig. 1. Word entropy (5) as a function of alphabet size Z, for $Z \leq 10$. Inset, left: $p(l)$ against word length l for alphabet sizes 2 to 5. Inset, right: $p(l)$ as a function of alphabet size for $l = 1, \ldots, 8$, where a small l gives a higher curve

3 Cellular Automata as a Function of ω

We now look at the impact of ω on the complexity of additive CA space-time output. We first introduce the concept of a reduced network.

Definition 1. *In cycle k, the reduced network is the set of affecting nodes $\mathcal{A}_i(k)$ of all nodes, together with the edges that connect affecting nodes $j \in \mathcal{A}_i(k)$ to their affected node i.*

For each counter k, we can draw up a reduced network. In [8], we show that this sequence of reduced networks asymptotically settles onto a fixed set of reduced networks. Formally, let us denote by $\mathcal{G}^r(k)$ the reduced network in cycle k. Then we obtain the sequence $\mathcal{G}^r(0), \mathcal{G}^r(1), \mathcal{G}^r(2), \ldots$ of reduced networks such that, for some $k \geq 0$, there exists $g \in \mathbb{N}$ such that $\mathcal{G}^r(k+g) = \mathcal{G}^r(k)$. The set $\mathcal{O} = \{\mathcal{G}^r(k), \mathcal{G}^r(k+1), \cdots, \mathcal{G}^r(k+g)\}$ is called a periodic orbit of reduced networks. This set is dependent on the initial set of update times $\boldsymbol{x}(0) = (x_1(0), x_2(0), \ldots, x_N(0))$ of the maxmin-ω system [8]. Figure 2 shows an example of such a sequence of reduced networks that enter a periodic orbit of size two; here the original network is a size 3 fully connected regular network (with neighbourhood size 3), whilst the system takes $\omega = 2/3$.

Pertinently, this means that, as $k \to \infty$, the maxmin-ω system can be replaced by a reduced system (with an underlying reduced network) with $\omega = 1$. This is intuitive - since only the affecting nodes affect the future state of a node i, it is equivalent to node i accepting all (such that $\omega = 1$) inputs from $\mathcal{A}_i(k)$.

It follows that $|\mathcal{A}_i(k)| \approx m$, where $m = \lceil \omega n \rceil$. This implies that the neighbourhood size of each node in $\mathcal{G}^r(k)$ should be approximately m. In fact, due to

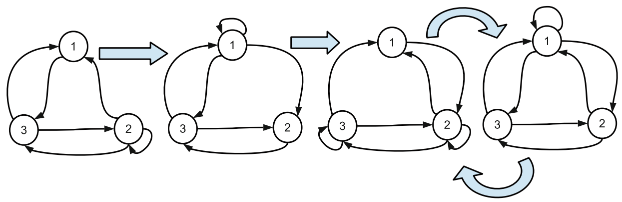


Fig. 2. Sequence of reduced networks of a maxmin-2/3 system where $N = 3$. Larger arrows indicate the transitions between successive iterations of the maxmin-2/3 system

the simultaneous arrival of some affecting nodes, $|\mathcal{A}_i(k)| \geq m$ [8]. Nevertheless, it is instructive to assume the average neighbourhood size of each node in a reduced network to be m (for example, see Fig. 2).

For further illustration, it is sufficient to take $g = 1$. Asymptotically then, we need only consider one underlying network; from here onwards, we shall take all mentions of "reduced network" to refer to this asymptotic reduced network, denoted $\mathcal{G}^r$. Thus, when m is small, the neighbourhood size of each node in $\mathcal{G}^r$ is small. This implies that some CA states – namely the fastest arriving ones – are favoured over other states. Consequently, the cellular automaton state space is narrower, giving fewer state possibilities and therefore a smaller Shannon entropy. Assuming that those CA states that appear with non-zero probability are equiprobable (due to the CA being additive), we can use (6) to say that, the Shannon entropy increases with the number of states. Thus, we obtain the following lemma.

Lemma 1. *The Shannon entropy of an additive CA pattern resulting from the maxmin-ω system is likely to increase with ω.*

We note that certain probability distributions of cell states actually correspond to a decrease in Shannon entropy with ω, e.g., if $p(s_1)$ is extremely large relative to other $p(s_j)$. However, this is contradictory to the property of additivity, which says that each CA state is approximately equiprobable. Thus, we assume that such extreme skewed probability distributions are rare so that Lemma 1 is almost always satisfied.

We move onto the analysis of word entropy as a function of ω. For ω small, the reduced network will have small neighbourhood size. Using the same arguments as earlier, the fastest CA states will prevail, giving few state possibilities. This is equivalent to having a small alphabet such that a range of word lengths are likely to be observed (see the inset of Fig. 1). For larger ω, the reduced network will have a larger neighbourhood size such that more CA states are prevalent; the likelihood of observing l-words is therefore decreasing with ω (again, see Fig. 1, inset). Thus, we expect word entropy to decrease with ω.

4 Experimental Results

We now run the maxmin-ω system and implement the additive cellular automaton rule (3). The underlying network is *regular*, equivalent to the one-dimensional CA lattice, and we take network size $N = 11$, where cells N and 1 are connected. We record the asymptotic values of Shannon and word entropies, along with the asymptotic quantities that summarise the update times of the maxmin-ω system itself. For this purpose, we require the following definitions.

Define the function $\mathcal{M}$ as the mapping $\mathcal{M} : \mathbb{R}^N \to \mathbb{R}^N$ whose components $\mathcal{M}_i$ are of the form of Eq. (1). We represent a system of N such equations by the following.

$$\boldsymbol{x}(k+1) = \mathcal{M}(\boldsymbol{x}(k)) \tag{9}$$

for $k \geq 0$, where $\boldsymbol{x}(k) = (x_1(k), x_2(k), \ldots, x_N(k))$. Denote by $\mathcal{M}^p(\boldsymbol{x})$ the action of applying $\mathcal{M}$ to a vector $\boldsymbol{x} \in \mathbb{R}^N$ a total of p times, i.e., $\mathcal{M}^p(\boldsymbol{x}) = \underbrace{\mathcal{M}(\mathcal{M}(\cdots(\mathcal{M}}_{p\text{ times}}(\boldsymbol{x}))\cdots))$.

Definition 2. *If it exists, the* cycletime vector *of* $\mathcal{M}$ *is* $\chi(\mathcal{M})$ *and is defined as* $\lim_{k\to\infty}(\mathcal{M}^k(\boldsymbol{x})/k)$.

Definition 3. *For some* $k \geq 0$*, consider the set of vectors*

$$\boldsymbol{x}(k), \boldsymbol{x}(k+1), \boldsymbol{x}(k+2), \ldots \in \mathbb{R}^N$$

where $\boldsymbol{x}(n) = \mathcal{M}^n(\boldsymbol{x}(0))$ *for all* $n \geq 0$*. The set* $x_i(k), x_i(k+1), x_i(k+2), \ldots$ *is called a* periodic regime *of* $i \in \mathbb{N}$ *if there exists* $\mu_i \in \mathbb{R}$ *and a finite number* $\rho_i, \in \mathbb{N}$ *such that* $x_i(k + \rho_i) = \mu_i + x_i(k)$*. The* period *of the regime is* ρ_i *and* $\chi_i = \mu_i/\rho_i$ *is the* cycletime of i*. The smallest* k *for which the periodic regime exists is called the* transient time*.*

Under our initial conditions, K_i will be finite (see [4], Theorem 12.7) and so, maxmin-ω always yields a periodic regime with the following system-wide quantities.

$$K = \max_i\{K_i\}, \quad \rho = \mathrm{LCM}_i(\rho_i), \quad \chi = (1/N)\sum_{i=1}^{N}\chi_i.$$

From now on, we take $\xi_i(k)$ and $\tau_i(k)$ to be independent of k, denoted ξ_i and τ_i, respectively. Our experiments can be described by the following steps.

1. Choose $\xi_i, \tau_i \in \mathbb{Z}$ both from the uniform distribution (with equal probability) taking largest value 5.
2. Choose an initial timing vector, $\boldsymbol{x}(0) = (0, \ldots, 0)$, and an initial CA state $\boldsymbol{s}(0)$ uniformly (with equal probability) from the alphabet $\Sigma = \{0, \ldots, Z-1\}$.
3. Iterate the maxmin-ω system 100 times for each ω value from 0.05 to 1, in steps of 0.05 (so there are 20 maxmin-ω systems to run).
4. For each maxmin-ω system, record the period ρ and cycletime χ, as well as the Shannon and word entropies.

5. Repeat above three steps 50 times to obtain, for each maxmin-ω system above, 50 independent periods and cycletimes, and Shannon and word entropies.
6. For each maxmin-ω system, record the mean of the 50 periods, cycletimes, Shannon and word obtained.

We vary the neighbourhood radius such that neighbourhood sizes explored are $n = 3, 5, 7, 9$, and 11. We also vary the alphabet size such that, for each n, the algorithm is run for alphabet sizes $|\Sigma| = 2, 3, \ldots, 10$. Figures 3 and 4 summarise the mean results.

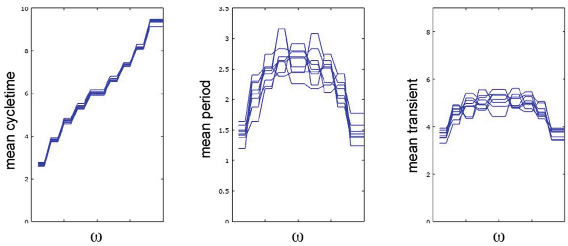


Fig. 3. Mean cycletime, period and transient time as a function of ω for a regular network of size 11 with neighbourhood size $n = 9$. Each curve represents each of the alphabet sizes 2 to 10. These graphs are typical for all other sizes $n = 3, 5, 7, 11$

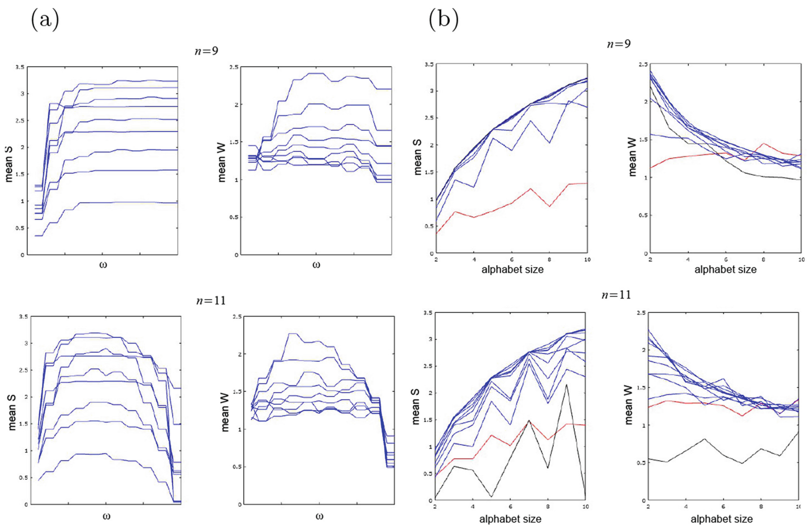


Fig. 4. Shannon S and word entropy W for a regular size 11 network with neighbourhood sizes $n = 9$ and 11 (results for $n = 3, 5, 7$ similar to $n = 9$). (a) Mean S and mean W versus ω. Each curve represents each alphabet size $Z = 2$ to 10. (b) Mean S and mean W versus Z. Each curve represents each $\omega \in \{0.05, 0.1, \ldots, 1\}$. Black: $\omega = 1$, blue: $0.1 \leq \omega \leq 0.95$, red: $\omega = 0.05$.

For all neighbourhood sizes except for $n = 11$, the Shannon entropy is an increasing value with ω (see Fig. 4(a)), in agreement with our analytical predictions. The word entropy behaves differently, however; it is increasing for most alphabet sizes, with a sharp decrease near $\omega = 1$. On the other hand, Fig. 4(b) agrees with our analytical predictions when alphabet size is approximately greater than or equal to 8; that is, W is a decreasing function of ω for a large alphabet size.

The logarithmic trend (6) of S with alphabet size is most apparent - this is attained when all CA states are equiprobable, and it is maximal, supported by the black Shannon entropy curve in Fig. 4(b), which is the $\omega = 1$ case. As expected, the word entropy is decreasing with alphabet size, although the case $\omega = 0.05$ is increasing (see red line in Fig. 4(b)).

The exception seems to be the case $n = 11$. Here, S is not increasing with ω, instead following a bell-like curve, taking maximal value at approximately $\omega = 0.5$. This Shannon entropy is minimal when $\omega = 1$ (see black/lowest curve in Fig. 4(b)), in extreme contrast to the logarithmic maximal trend for other neighbourhood sizes.

We end this section by combining the Shannon and word entropy results into one (S, W)-plane. Thus, for each n, consider a fixed alphabet size. This gives a value (mean S, mean W) for each ω value. To find which of these produces the 'most complex' point, we take the following simple distance from the origin, for $\omega \in (0, 1]$.

$$d_\omega = \sqrt{\bar{S}^2 + \bar{W}^2} \tag{10}$$

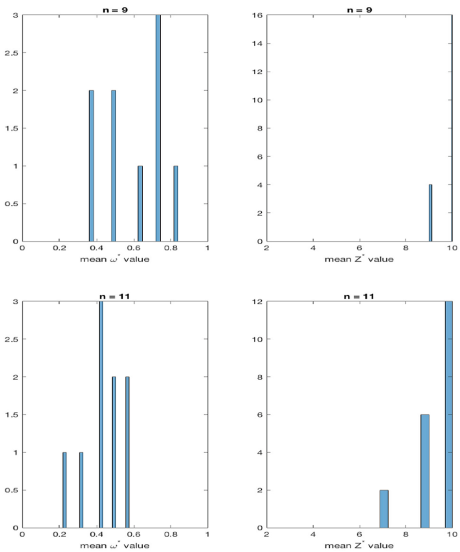


Fig. 5. Frequency of ω^* and Z^* values that yield the most complex CA patterns. Left: Frequency of ω^* values. Right: Frequency of Z^* values. We present results for neighbourhood sizes $n = 9$ and $n = 11$ as they show the most variation

where $\bar{S}$ = mean S and $\bar{W}$ = mean W. Then, we are interested in $\omega^* = \arg\max_\omega d_\omega$, i.e., the ω values that maximise d_ω. Often there exist more than one such ω. For example, for $n = 3$, $\omega^* = \{0.75, 0.8, 0.85, 0.9, 1\}$ for all alphabet sizes. In such cases, we take the mean of this list of ω^* values.

Thus, for each alphabet size, we have one ω^* value. Performing this calculation for all alphabet sizes up to 10, we produce histograms indicating where such ω^* values congregate; these are shown in Fig. 5. A similar calculation finds, for a fixed ω value, the alphabet size that produces the 'most complex' (S, W) point; we denote this alphabet size Z^*, also depicted in Fig. 5.

5 Discussion

We have demonstrated the effect of maxmin-ω dynamics on an additive cellular automaton in two orthogonal ways: asynchrony was imposed via maxmin-ω and, secondly, the alphabet of state possibilities was extended.

We previously noted some correspondence in complexity between timing and CA pattern when the alphabet was binary [8,9]. Here, we have shown that an larger alphabet generates additional facets to the story of complexity. Thus, we claim that the essence of the maxmin-ω system is best captured by a CA with a larger number of states than two. Whilst complexity does not follow simple bell-like curves, the (S, W)-plane offers some support to the argument that the most complex patterns occur when $\omega \approx 0.5$ (see Fig. 5).

It must be mentioned that the entropic measure of complexity employed in this paper is just one of various probabilistic measures that are in use. The measures that immediately seem relevant to our work are the LMC complexity [5] and the Rényi entropy [10]. An in-depth exploration is saved for a future study, suffice it to say that calculation of such measures is straightforward.

Let us briefly discuss the LMC complexity C. For our model, we construct two of these measures: (i) $C_S = SD_S$, where $D_S = \Sigma_{j=1}^{Z} (p(s_j) - 1/Z)^2$ is a measure of 'disequilibrium'; (ii) $C_W = WD_W$, where D_W is a measure of disequilibrium, given by $D_W = \Sigma_{l=1}^{T} (p(l) - 1/T)^2$. The values $1/Z$ and $1/T$ respectively denote the probability of observing a CA state and an l-word if all states and all words are equiprobable. Since our additive rules imply $p(s_j) = 1/Z$ for all j, we have $D_S = 0$ so that $C_S = 0$. To calculate C_W we substitute (8) in D_W to obtain a decreasing C_W with alphabet size. Thus, both C_S and C_W tend to behave like ideal gases, where a variety of CA patterns are observed.

The LMC complexity as a function of ω is more interesting. We predict that, for small ω, since only a few cell states prevail (the fastest ones), then the overall behaviour is akin to a crystal - predictably periodic; as ω increases, more CA states become admissable, giving a wider probability distribution, though each non-zero probability is equal. Thus, D_S is decreasing with ω and, since S is increasing with ω, we have an intriguing situation to that of Fig. 1 of [5], with C_S maximising for some value of ω between 0 and 1, and disappearing when $\omega = 0$ and 1. Since $p(l)$ is expected to decrease with ω, we have $D_W \to 0$ as ω increases. Therefore, $C_W \to 0$ as $\omega \to 1$.

In all cases, although the additive rule is considered to be in Wolfram classes I and II [14], this CA is shown to produce a variety of space-time patterns, depending on ω, and particularly when the alphabet is enlarged; thus, all four Wolfram classes may be exhibited simply by controlling ω and $|\Sigma|$.

References

1. Fatès, N.A., Morvan, M.: An experimental study of robustness to asynchronism for elementary cellular automata. Complex Syst. **16**, 1–27 (2005)
2. Granovetter, M.: Threshold models of collective behavior. Am. J. Sociol. **83**(6), 1420–1443 (1978)
3. Gutowitz, H., Langton, C.: Mean field theory of the edge of chaos. In: European Conference on Artificial Life, pp. 52–64. Springer, Heidelberg (1995)
4. Heidergott, B., Olsder, G.J., Van der Woude, J.: Max Plus at Work: Modeling and Analysis of Synchronized Systems: A Course on Max-Plus Algebra and Its Applications. Princeton University Press, Princeton (2006)
5. Lopez-Ruiz, R., Mancini, H.L., Calbet, X.: A statistical measure of complexity. Phys. Lett. A **209**(5–6), 321–326 (1995)
6. Marr, C., Hütt, M.T.: Topology regulates pattern formation capacity of binary cellular automata on graphs. Phys. A Stat. Mech. Appl. **354**, 641–662 (2005)
7. McCulloch, W.S., Pitts, W.: A logical calculus of the ideas immanent in nervous activity. Bull. Math. Biophys. **5**(4), 115–133 (1943)
8. Patel, E.L.: Maxmin-plus models of asynchronous computation. Ph.D. thesis, University of Manchester (2012)
9. Patel, E.L.: Maxmin-ω: a simple deterministic asynchronous cellular automaton scheme. In: International Conference on Cellular Automata, pp. 192–198. Springer, Cham (2016)
10. Rényi, A.: On measures of entropy and information. Technical report, Hungarian Academy of Sciences, Budapest (1961)
11. Schönfisch, B., de Roos, A.: Synchronous and asynchronous updating in cellular automata. BioSystems **51**(3), 123–143 (1999)
12. Watts, D.J.: A simple model of global cascades on random networks. Proc. Natl. Acad. Sci. **99**(9), 5766–5771 (2002)
13. Wolfram, S.: Statistical mechanics of cellular automata. Rev. Mod. Phys. **55**, 601–644 (1983)
14. Wolfram, S.: Universality and complexity in cellular automata. Phys. D Nonlinear Phenom. **10**(1–2), 1–35 (1984)

Evolutionary Development and the VCRIS Model of Natural Selection

John M. Smart(✉)

Naval Postgraduate School, Monterey, CA 93943, USA
johnsmart@gmail.com

Abstract. This paper offers a systems definition of the phrase evolutionary development (evo devo, ED), and a few examples of generic evolutionary and developmental process in autopoetic (self-reproducing) systems. It introduces a toy conceptual model, the VCRIS evo devo model of natural selection, exploring autopoetic selection in both evolutionary and developmental terms. It includes an empirical observation, the 95/5 rule, generalized from observations in evo-devo biology, to offer a preliminary sketch of the dynamical interaction of evolutionary and developmental processes in living replicators. Autopoetic models may be applied to both to living systems and to nonliving adaptive replicators at many scales, even to the universe as a complex system, if it is a replicator in the multiverse. Evo devo models offer potentially fundamental dynamical and informational ways to understand autopoetic systems. If such models are to become validated in living systems, and generalized to nonliving autopoetic systems, they will require significant advances in both simulation and theory in coming years.

Keywords: 95/5 rule · Autopoetic systems · Biological development Complex adaptive systems · Convergent evolution · Developmental selection Evo-devo biology · Evo devo models · Evolutionary biology Evolutionary development · Evolutionary selection · Fine-tuning Multiverse · Natural selection · Niche construction Modern evolutionary synthesis · Self-Organization · Stigmergy Stochastic processes

1 Definition and Overview

In both our modern evolutionary synthesis during the 20th century (Huxley 1942), and in new proposals for an 'extended evolutionary synthesis' (Pigliucci 2007) evolutionary process, while now acknowledged to be multi-causal, is still considered grossly unpredictable in its long-term complex dynamics, since on Earth at least, collective environmental diversity and contingency all appear to grow robustly with time. Evolutionary theorist Steven Jay Gould famously captured this view in a thought experiment. He proposed that if we were to view the "tape of life" on another Earthlike planet, virtually none of life's species, functions and morphologies would turn out the same as on our Earth (Gould 1977, 2002). Beginning in the 1990s a few modern evolutionary scholars, most notably Simon Conway-Morris, began to argue the opposite, proposing

A. J. Morales et al. (Eds.): ICCS 2018, SPCOM, pp. 108–119, 2018.
https://doi.org/10.1007/978-3-319-96661-8_11

that many of life's functions, morphologies, and species types would predictably arise in a potentially wide range of Earth-like environments (Conway-Morris 1998, 2004, 2015). Scholars of convergent evolution, which includes genetic, phenotype and population models, are a leading academic community proposing that implicit predictability exists in a subset of evolutionary processes (Müller and Newman 2003; Ogura et al. 2004; Vermeij 2006; Dryden et al. 2008; McGhee 2011; Powell 2012; Flores-Martinez 2014; McLeish 2015; Powell and Mariscal 2015; Losos 2017). In an effort to better systematize these two opposing perspectives, this paper offers a phrase, evolutionary development, and a simple conceptual model, the VCRIS model of natural selection, and argues that each must be evaluated within a framework of autopoetic systems.

Autopoesis is a term introduced by Chilean biologists Maturana and Varela (1973/1980) to describe the chemistry of living cells. It became popular with a few systems theorists in the late 20th century to describe the capacity of some complex systems to self-reproduce and self-maintain. Autopoesis scholars seek to find general systems rules applicable to any stably self-reproducing complex systems, including not only living systems, but stars, the chemical origin of life, and ideas, behaviors, algorithms, organizational rulesets, and technologies in culture. Implicit to autopoetic systems models is the idea that a better information theory, including a theory of cumulative adaptive intelligence in the replicator, its inheritance system, and its environment, will be necessary to understand dynamical change in complex systems (Maturana and Varela 1987; Mingers 1995; Luisi 2003; Bourgine and Stewart 2004).

For any potentially autopoetic system, we can propose a useful descriptive phrase, "evolutionary development", "evo devo" or "ED" as a replacement for the more general term "evolution", to describe that system in this language of opposing dynamical perspectives. It is a useful phrase to employ whenever any scholar thinks that both experimental, creative, contingent, stochastic, and increasingly unpredictable or "evolutionary" processes, and conservative, convergent, statistically deterministic (probabilistically predictable) or "developmental" processes, including a replicative life cycle, may be each contributing to selection and adaptation in any complex system. Use of this oppositional term also communicates our humility and ignorance when we are asked whether evolutionary (divergent) or developmental (convergent) process are presently dominating in any system or environment. We usually don't know which processes are most in control of physical or informational dynamics, at first glance. Careful study, modeling, and data collection are often required to see where any complex system is presently headed, process by process.

The hyphenated "evo-devo" is commonly used for living systems, most prominently in evo-devo genetics and epigenetics, where this oppositional perspective is most rigorously being explored today (see Müller and Newman 2003; Schlosser and Wagner 2004; Carroll 2005; Callebaut and Rasskin-Gutman 2005; Pigliucci and Müller 2010) and the unhyphenated "evo devo" can be used for the theory of any potentially replicating and adapting complex system (star, prebiotic system, gene, cell, organism, meme (concept), behavior, technology), whether living or nonliving. Inspired by the work of evo-devo biologists, evo devo systems theorists look for processes of evolutionary divergence and experiment and developmental convergence and constraint working simultaneously in any self-reproducing complex systems, at any scale.

Evo devo systems theory thus redefines the term "evolution", within any autopoetic system, to restrict evolutionary process to stochastic, information-creative, experimental, diversifying, and nonhierarchical processes of system change. These processes are the dynamical and informational opposite of the predictable, information-conservative, convergent, unifying, and hierarchical processes of "development." Redefinitions are never a popular choice, but this redefinition is potentially clarifying for autopoetic dynamics, from the perspective of information theory. If evolutionary processes necessarily generate new information, and developmental processes conserve and build upon old information, and we can determine "new" or "old" only in relation to the life cycle of the system under analysis, we may have a useful new perspective on both dynamics and their intrinsic predictability.

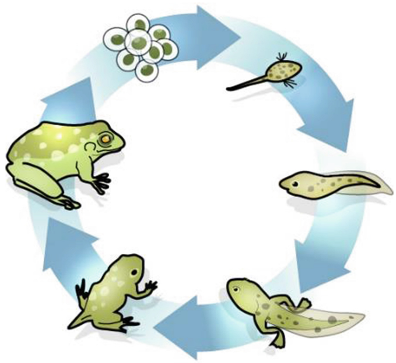


Fig. 1. An evo devo life cycle. Evo-devo biology offers us many empirically observable ways in which evolutionary process creates new information, and developmental process conserves old information, with respect to the prior autopoetic cycle.

When we apply these definitions to the life cycle of an individual organism, as in Fig. 1, we can observe evolutionary, information-creative processes in such events as stochastic gamete production, and in the stochastic cellular microarchitecture in any specific frog. Simultaneously, we can observe developmental, information-conservative processes in any replicative dynamics, informatics, and morphology that we empirically observe in all frogs of a specific species. Both evolutionary and developmental processes can thus be empirically differentiated in any living complex system via these definitions. Both processes are presumably fundamental to adaptation, and the ways each system encodes representations (models, intelligence) of itself and its environment.

Table 1. Polar word pairs with homology to evolution and development as each pair are commonly modeled in biological systems.

Evolution	Development
Unpredictability	Predictability
Chance	Necessity
Stochastic	Destined
Divergent	Convergent
Reversible	Irreversible
Possibilities	Constraints
Uniqueness	Sameness
Transformation	Conservation
Accidental	Self-Organizing

(*continued*)

Table 1. (*continued*)

Evolution	Development
Unpredictability	Predictability
Bottom-Up	Top-Down
Local	Global
Immature	Mature
Individual	Collective
Instance	Average
Short-Term	Long-Term
Creativity (Novelty)	Discovery (Universality)
Period-Doubling (Chaos)	Period-Halving (Order)
Experimental	Optimal
Neoteny	Differentiation
Belief (unproven)	Knowledge (verified)
Innovation	**Sustainability**

2 Two Polar Categories and Tensions

Table 1 (Smart 2008) introduces sets of two polar (equal and opposite) word pairs that can be associated with evolutionary and developmental processes in a range of complex systems. As you look them over, think of all the events, processes, and systems you have previously described with these words. These and similar words, and the concepts behind them, are often useful starts at categorizing social, economic, and technological events and processes into one of two camps.

Some systemic processes operate by chance, others by necessity, and some by both. Some processes are random, others predestined. Some events are indeterminate, others predetermined. Some processes are segregating, others integrating. Some act bottom-up, others are top-down. Some systems appear to be branching, others funneling. Some changes look reversible, others irreversible. Some are generating novelty, others conserving sameness. Some are exploring possibilities, others running into constraints. Some promote variability, others stability. Some degrade hierarchies, others create hierarchy. In the organization, good foresight and strategy requires a continual balance between divergent (innovative, experimental) and convergent (predictive, conservative) thinking. We can see these twin tensions, and their mixture, in all the ways humans use for knowing the world.

In the twentieth century, we learned that even our scientific laws fall neatly into these two categories. From our reference frame, we have discovered deterministic (developmental) types of laws that precisely describe the far future, like the equations of classical mechanics and relativity. We have also discovered stochastic and statistical (evolutionary) physical laws, like quantum mechanics, thermodynamics, and nuclear physics.

We have also learned we can view physical and informational systems as either deterministic or stochastic, depending on the analytical reference frame we adopt.

Deterministic laws are highly conserved and predictable at the individual level (i.e., the laws of motion for individual objects), yet become unpredictable at the collective/emergent level (i.e., the N-body problem in physics). Stochastic laws are random, novel, and creative at the individual level (the quantum state or entropy of any particular system, the decay of any particular nucleus), and yet are probabilistically predictable at the collective level. We see a simple example in radioactive half-life, and more complex examples in non-equilibrium thermodynamics, self-organized criticality, and phase transition thermodynamics. Such factors as the reference frame of the observer with respect to the system, the scale at which they are observing the system, and the duration of observation relative to a (presumed) autopoetic cycle all seem to influence the ease and extent of predictability in nature.

Many social, economic, and political processes historically alternate between unpredictable and divergent (evolutionary) and predictable and convergent (developmental) phases (cf. Vermeij 2009). For every social issue, we can find processes simultaneously generating "evolutionary" variety and "developmental" convergence, in comparative analyses of different cities, counties, states, countries, or regions. For example, regarding economic inequality, we find great "evolutionary" variations, country by country, in the levels and quality of social services available to each citizen, and in the cycles of increasing or decreasing inequality. Yet we also find a long-term "developmental" trend of predictably increasing total economic inequality (relative and absolute rich-poor divides) the greater the flow rates of capital, goods, and information in any societies we analyze (Bejan and Errera 2017). The two opposing perspectives and tensions of evolution and development (unpredictability or predictability) appear to be equally fundamentally useful ways to view the world.

3 The VCRIS Model of Natural Selection in Autopoetic Systems

If we wish to understand natural selection in autopoetic systems, both living and non-living, we must better characterize dynamical change, and develop better theories of information and intelligence. The VCRIS ("vee-kriss") evo devo conceptual model (Smart 2017) may be a useful, small step toward these challenges, especially when contrasted to the classic VIST model (variation, inheritance, selection, time/cumulative replication, Russell 2006) of dynamics offered by traditional evolutionary theory (Fig. 2).

The VCRIS model proposes that three sets of physical and informational dynamics must be modeled to understand and predict the outcomes of natural selection in autopoetic systems. The first two are fundamentally oppositional processes, and the third arises from their interaction. These are:

1. Variational or "Evolutionary" processes that generate, maintain and manage diversity, divergence, and experiment. When we observe them from within any autopoetic system, these processes grow increasingly unpredictable over time.
2. Convergent or "Developmental" processes that attract, constrain, maintain and guide the system through hierarchical stages of form and function. When we observe them from within any autopoetic system, these processes grow increasingly predictable over time.

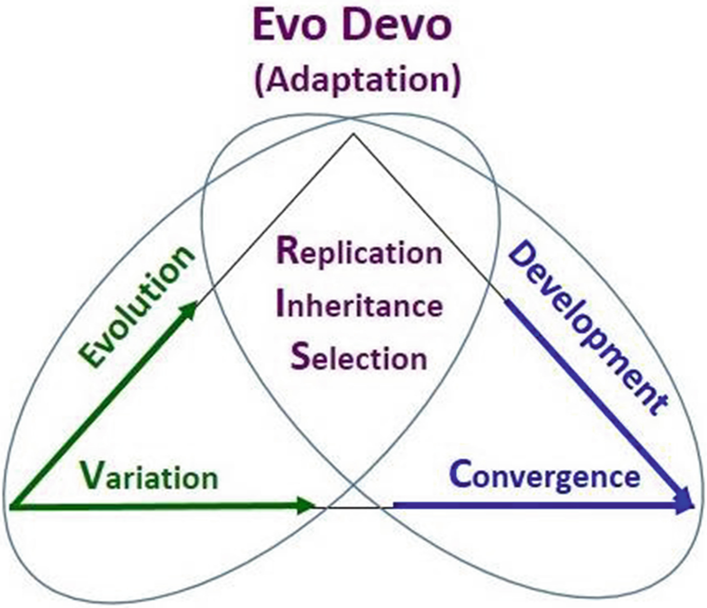


Fig. 2. The VCRIS toy model of natural selection in autopoetic systems. Variation (evolutionary process) and Convergence (development), operating under Replication and Inheritance, can be viewed as fundamental physical and informational dynamics that mediate Selection in self-reproducing systems.

3. "Evo Devo" processes that are successfully replicative, allow inheritance, and are adaptive. These processes are always some blend of the first two fundamental types. In the VCRIS model, adaptive processes can be further divided into Replication (Organism) processes, Inheritance (Seed, Genes) processes, and Selection (Environment) processes.

In the VCRIS model, physical and informational processes that change unpredictably in successive replication cycles, to generate, maintain, and manage Variety, are in tension and opposition with physical and informational processes that change predictably in successive replication cycles, and thus generate, maintain, and manage Convergence. V and C are the first two terms in the VCRIS model, as these two oppositional processes are proposed as "root perspectives" in any model of physical and informational change in autopoetic systems, including our universe itself, if it is a replicating and adaptive system, as various theorists have proposed (Smolin 1992, 1997, 2004; Vaas 1998; Vidal 2010; Price 2017). Standard evolutionary theory offers no model of this fundamental opposition, at every scale at which replication occurs, including gene, epigene, organism, group, niche, environment, and universe.

In toy cellular automata models of the universe, like Conway's Game of Life (Poundstone 1985), the spatiotemporally repetitive structures and dynamics that we see in each successive game (replication cycle), can be defined as predictable, convergent and developmental. Such reliably emergent structures and dynamics are robust to variation in most of the game's initial conditions (occupied configurations within the initiating matrix), yet they are also finely sensitive (finely tuned) to be critically dependent on a few of those conditions, such as the rules of the automata. The morphology and dynamics of other emergent structures in this game are essentially unpredictable, divergent, and can be thought of as sources of evolutionary variety within the game. See Poundstone for an account of Conway's game from a universal perspective.

In more complex models, like the replication and adaptation of individual living organisms, the features of two genetically identical twins that look the same are (in theory) predictable, convergent, and developmental. The morphological, dynamical and functional features that are stochastically different, which include their fingerprints, brain wiring, organ microarchitecture, and many (not all) of their ideas and behaviors, are unpredictable, variety-generating (within bounds), and "evolutionary," in our potentially more precise and useful autopoetic use of this term.

Perhaps the next most fundamental concepts we must employ to understand selection and change in autopoetic systems are the ideas of complexity and system themselves, and the conditions under which these systems emerge, develop boundaries, maintain themselves, encode models of self- and environment, and become adaptive. Together with Variation and Convergence, Replication, or the cyclic maintenance and copying of a complex system in space-time, and the Inheritance of parameters that influence that replication, give us the VCRI terms in the VCRIS model.

In this model, natural selection involves "tree-like" evolutionary processes driven by Variation (creation of new information) and "funnel-like" developmental processes driven by Convergence (conservation of old information). These evo devo processes act simultaneously, and often in opposition to each other, in service to adaptation. Consider how all Replicating organisms are sometimes driven to variation, and sometimes to convergence. Inheritance units (seeds, genes) sometimes duplicate (think of gene duplication) and vary, and sometimes converge (with gene loss). Selection in the environment sometimes favors diversity, and sometimes favors a particular phenotype. In the VCRIS model, evo and devo (variety-creating and convergent) replication and inheritance under selection are the root source of adapted order.

Along with Selection itself, VCRI processes offer us a complex systems framework for such emergent properties as persistence and stability under stress, continual and periodic system renewal, growth, and the possibility for improvement (complexification, learning, intelligence) in complex systems over time.

We may also use the evo devo perspective to gain a new perspective on another long-used term in the complexity literature, self-organization. In autopoetic systems, a hidden developmental dynamic will make any system appear to be "self-organizing". Cut up a virus in a petri dish, and its molecular fragments will "self-organize" back into virus. They do so because those molecules have adaptively fine-tuned, over many prior life cycles, to use physically and informationally metastable features of the universal environment. Self-organization, then, may be at least partially a process of hidden development, finely tuned, via top-down causality, in previously stable autopoetic cycles.

The autopoetic framework also requires us to say more about the interaction of intelligence and the environment. The more intelligent the organism, the more they can "niche construct" their local environment to make it more co-adapted (Odling-Smee et al. 2003). Alternatively, some scholars call this process "stigmergy" (Heylighen 2008, 2016). Niche construction/stigmergy is an underrated informational process that clearly exerts top-down causality and developmental constraint, in ways we haven't yet characterized well in complexity theory.

In living autopoetic systems, historically metastable features of the local environment are used by genes to reliably guide the evolving and developing organism to its future destinations. Environments may also replicate, on some higher systems level, just as organisms and seeds replicate. This happens, for example, when we replicate an urban architecture or idea-complex (like capitalism or democracy), when stars replicate, when continents drift apart, or if our universe itself replicates.

If we live in an autopoetic (evo devo) universe, our selective environment can itself be thought of as at least partly analogous to an "organism," a system fated to produce a new seed or seeds in special high-complexity locales. This is a view that is currently quite rarely discussed in complexity research. In addition, adapted intelligence in any evo devo system is always opportunistically partitioned between three complex actors, Seed, Organism, and Environment (SOE partitioning). Intelligence is never resident in only one of these actors. It always seems to straddle all three (Smart 2008).

In my view, modern evolutionary biology offers a systems view of life and selection that is dangerously incomplete. It has long neglected the physical and informational roles of development, and the (admittedly speculative) idea that our universe (environment) itself may be developing. Fortunately, evo-devo biology is helping to rehabilitate development as a process in living systems. We must also better characterize development at societal, global, and universal scales. I am hopeful for the value of this view in improving our models of human civilization, which seems to have many still poorly characterized developmental processes, which in turn must broadly influence selection.

Biologically-Inspired Complexity Science and Philosophy (BICS&P) might be reasonable general title for this still-emerging field of evo devo systems and complexity theory. BICS&P is the self-description of our Evo Devo Universe academic discussion community. (EvoDevoUniverse.com), where the author and a small community of interdisciplinary scholars in cosmology, physics, astrobiology, chemistry, biology, social science, computer science, the complexity sciences, philosophy, and other disciplines have explored various versions of these ideas since 2008.

4 Evolutionary Development in Organisms: The 95/5 Rule

Since the mid-1990s, the interdisciplinary subfield of evolutionary developmental, or "evo-devo" biology has emerged to explore the relationship between evolutionary and developmental processes at the scale levels of single-celled and multicellular organisms (Steele 1981; Jablonka and Lamb 1995; Raff 1996; Arthur 2000; Wilkins 2001; Hall 2003; Müller and Newman 2003; Verhulst 2003; West-Eberhard 2003; Schlosser and Wagner 2004; Carroll 2005; Callebaut and Rasskin-Gutman 2005).

For many authors in this subfield, the fundamental role of evolutionary processes can be hypothesized as cumulative mechanisms that generate experimental ("good bet") types of diversity, to improve the odds of survival under environmental selection. Evolutionary systems use stochastic processes in an increasingly information-driven and intelligent way as organic complexity grows, but evolutionary innovation itself is largely unpredictable (Shapiro 2011; Noble 2017). Living systems continually sense their internal states and environment, and they react to catastrophe and stress with bursts of such poorly-predictable, information-driven innovation, a pattern some evolutionary biologists call punctuated equilibrium (Eldredge and Gould 1972).

The fundamental role of developmental processes in evo-devo biology can be hypothesized as cumulative mechanisms that conserve and execute a small subset of (in-principle) predictable processes that have worked in the past to guarantee replication, under a range of chaotic internal and external environmental conditions. Developmental systems encode future-predictive probabilistic models of themselves and their environment, models which we assume follow the rules of Bayesian probability in nervous systems and presumably even in single celled organisms. Developmental prediction (a convergent form of "intelligence") is generated from special initial conditions (developmental genes), tuned via informational constancies that exist in genes, developing organisms, and the environment.

The theory of facilitated variation (Gerhart and Kirschner 2005; 2007), in which the genetic processes in living systems are assumed to sort into two groups, a conserved core, which regulate critical elements of development and physiology, and a set of changing genetic elements, whose variation is "facilitated" by the conserved core, presumably in ways that reduce the lethality of experimental change and increase the utility of genetic "experiments" that are retained by populations, is a model consistent with this view. In evo devo language, the conserved core are conserved developmental genetic, allelic, and epigenetic processes, and evolutionary genetic processes are those that facilitate genetic, allelic, and epigenetic variation within and across generations. Such processes presumably act in tension with and opposition to each other in very fundamental informational and dynamical ways.

In this model, natural selection can be argued to be a composite of two more fundamental kinds of selection. Evolutionary selection biases the system toward potentially useful, intelligence-guided innovation and disorder when needed, and developmental selection biases the system toward convergences and order that have historically allowed complexity conservation and replication. In this view, we must see both of these selective and often opposing processes, apparently at work at many scales in every system that replicates, to truly understand biological change. For example, we should be able to identify both structure- or function-divergent and structure- or function-convergent gene flow operating between species in the terrestrial biosphere, via such processes as genetic drift and horizontal (or lateral) gene transfer. Such transfer is well documented in Prokarya, and is greatly facilitated by viruses in Eukarya (Zimmer 2015).

One of the clarifying features of developmental selection is that it is always critically dependent on a small subset of control parameters (in biology, genes and other regulatory molecules). While about half of metazoan genes are expressed in such processes as organ development, less than 20% of these (thus less than 10% of all genes) are substantially regulated during expression (Yi 2010). A further subset of our genome, roughly 5% of DNA in human, mouse, and rat, is highly conserved across these and other metazoan species. This 5% of our genome typically cannot be changed without stopping, or causing major deleterious effects to, processes of development. The majority of this highly conserved DNA, 3.5% of our genome, is noncoding, yet presumably also constrains functional expression (Wagman and Stephens 2004). A subset of this conserved DNA is sometimes called the developmental genetic toolkit (DGT), or less accurately, the evo-devo gene toolkit. These genes include the *Hox* genes which determine animal body plans, and they often involve initial symmetry breaking choices in spatial, dynamical, and informational form and function that commit the organism to a particular developmental path. Thus a subset of all metazoan genomes have become very finely tuned, over many past replications, for the production of complex, path dependent modularity, hierarchy and life cycle. Presumably, the other 95% of these genomes can change and generate diversity without such immediately deleterious effects.

Thus all genomes can be categorized into two groups, of conserved and non-conserved genes, and we can propose that all highly conserved genes which are *also* highly tuned (highly sensitive to change, with deleterious effect in any autopoetic cycle) are the core constraints on development itself as a dynamical process. I call this observation the 95/5 rule, and have found early evidence for it in replicative systems at a wide variety of scales (Smart 2008). The rule proposes that a small subset of developmental parameters are top-down causal, involving one-way information flow (in this case, developmental genes to organism). They can no longer be easily changed, they can only be added to, as organisms get more complex. The remainder of the genome can be considered evolutionary, whether it controls evolutionary or developmental process, as all of those genes can be altered by two-way information flow with the environment, with feedback. Per the 95/5 rule, a small, and highly-tuned set of top-down causal parameters (constraints) may always be required to exist, in any evo devo system.

Evo-devo genetics and epigenetics are rapidly improving fields, and they may lead us to other potentially general rules for autopoetic systems. I am hopeful that a modern, evo devo approach to autopoetic systems may help us better understand, within such systems, a number of presently poorly defined concepts as top-down and bottom-up causality, adaptation, intelligence, self-organization, and convergent evolution.

Acknowledgments. The author thanks Evo Devo Universe co-directors Clement Vidal, Georgi Georgiev, Michael Price, and Claudio Flores-Martinez, and the 100 or so scholars on the EDU-Talk listserve for helpful discussions. Thanks also to Carlos Gershenson and the CCS2018 committee for approving our satellite meeting on Evolution, Development, and Complexity at CCS2018, where earlier versions of these ideas were discussed.

References

Arthur, W.: The Origin of Animal Body Plans. Cambridge U. Press, Cambridge (2000)
Bejan, A., Errera, M.R.: Wealth inequality. J. Appl. Phys. **121**(12), 124903 (2017)
Bourgine, P., Stewart, J.: Autopoiesis and Cognition. Artif. Life **10**, 327–345 (2004)

Callebaut, W., Rasskin-Gutman, D.: Modularity: Understanding the Development and Evolution of Natural Complex Systems. MIT Press, Cambridge (2005)
Carroll, S.B.: Endless Forms Most Beautiful. Norton, New York (2005)
Conway-Morris, S.: The Crucible of Creation: The Burgess Shale and the Rise of Animals. Oxford U. Press, Oxford (1998)
Conway-Morris, S.: Life's Solution: Inevitable Humans in a Lonely Universe. Cambridge U. Press, Cambridge (2004)
Conway-Morris, S.: The Runes of Evolution: How the Universe Became Self-Aware. Templeton Press, West Conshohocken (2015)
Dryden, D.T.F., Thomson, A.R., White, J.H.: How much of protein sequence space has been explored by life on Earth? J. R. Soc. Interface **5**, 953–956 (2008)
Eldredge, N., Gould, S.J.: Punctuated equilibria: an alternative to phyletic gradualism. In: Schopf, T.M. (ed.) Models in Paleobiology, Freeman & Cooper (1972)
Flores-Martinez, C.L.: SETI in the light of cosmic convergent evolution. Acta Astronaut. **104**(1), 341–349 (2014)
Gerhart, J.C., Kirschner, M.W.: The Plausibility of Life. Yale U. Press, New Haven (2005)
Gerhart, J.C., Kirschner, M.W.: The theory of facilitated variation. PNAS **104**(Suppl), 8582–8589 (2007)
Gould, S.J.: Ontogeny and Phylogeny. Harvard U. Press, Cambridge (1977)
Gould, S.J.: The Structure of Evolutionary Theory. Harvard U. Press, Cambridge (2002)
Hall, B.K. (ed.): Environment, Development, and Evolution. MIT Press, Cambridge (2003)
Heylighen, F.: Accelerating socio-technological evolution: from ephemeralization and stigmergy to the Global Brain. In: Modelski, G., Devezas, T., Thompson, W.R. (eds.) Globalization as Evolutionary Process, Routledge (2008)
Heylighen, F.: Stigmergy as a Universal Coordination Mechanism. Cogn. Syst. Res. **38**, 4–13 (2016)
Huxley, J.: Evolution: The Modern Synthesis. George Allen & Unwin, London (1942)
Jablonka, E., Lamb, M.J.: Epigenetic Inheritance and Evolution: The Lamarckian Dimension. Oxford U. Press, Oxford (1995)
Losos, Johnathan B.: Improbable Destinies: Fate, Chance, and the Future of Evolution, Riverhead Books (2017)
Luisi, P.L.: Autopoiesis: a review and a reappraisal. Naturwissenschaften **90**, 49–59 (2003)
Maturana, H.R., Varela, F.J.: Autopoiesis and Cognition: The Realization of the Living. D. Reidel, Dordrecht (1973/1980)
Maturana, H.R., Varela, F.J.: The Tree of Knowledge: The Biological Roots of Human Understanding. Shambhala, Boston (1987)
McGhee, G.R.: Convergent Evolution: Limited Forms Most Beautiful. MIT Press, Cambridge (2011)
McLeish, T.C.B.: Are there ergodic limits to evolution? Ergodic exploration of genome space and convergence. Interface Focus **5**, 41–53 (2015)
Mingers, J.: Self-Producing Systems. Kluwer Academic/Plenum, New York (1995)
Müller, G.B., Newman, S.A. (eds.): The Origination of Organismal Form: Beyond the Gene in Developmental and Evolutionary Biology. MIT Press, Cambridge (2003)
Noble, D.: Evolution viewed from physics, physiology and medicine. Interface Focus **7**(5), 20160159 (2017)
Odling-Smee, J., et al.: Niche Construction: The Neglected Process in Evolution. Princeton U. Press, Princeton (2003)
Ogura, A., Kazuho, I., Gojobori, T.: Comparative analysis of gene expression for convergent evolution of camera eye between octopus and human. Genome Res. **14**, 1555–1561 (2004)
Pigliucci, M.: Do we need an extended evolutionary synthesis? Evolution **61**, 2743–2749 (2007)

Pigliucci, M., Müller, G.B. (eds.): Evolution: The Extended Synthesis. MIT Press, Cambridge (2010)
Poundstone, W.: The Recursive Universe: Cosmic Complexity and the Limits of Scientific Knowledge. William Morrow & Co, New York (1985)
Powell, R.: Convergent evolution and the limits of natural selection. Euro. J. Phil. Sci. **2**, 355–373 (2012)
Powell, R., Mariscal, C.: Convergent evolution as natural experiment: the tape of life reconsidered. Interface Focus **5**, 40–53 (2015)
Price, M.: Entropy and selection: Life as an adaptation for universe replication, Complexity (2017)
Raff, R.: The Shape of Life: Genes, Development, and the Evolution of Animal Form. U. of Chicago Press, Chicago (1996)
Russell, C.M. (2006). Epicofevolution.com/biological-evolution. Accessed 5 Mar 2018
Schlosser, G., Wagner, G.P. (eds.): Modularity in Development and Evolution. U. of Chicago Press, Chicago (2004)
Shapiro, J.A.: Evolution: A View from the 21st Century. FT Press Science, Upper Saddle River (2011)
Smart, J.M.: Evo devo universe? a framework for speculations on cosmic culture. In: Steven, J. D., Mark, L.L. (eds.) Cosmos and Culture, NASA Press (2008)
Smart, J.M.: Evo devo foresight. In: The Foresight Guide (2017). Foresightguide.com/evo-devo-foresight-table-of-contents. Accessed 28 Mar 2018
Smolin, L.: Did the Universe Evolve? Class. Quantum Gravity **9**, 173–191 (1992)
Smolin, L.: The Life of the Cosmos. Oxford U. Press, Oxford (1997)
Smolin, L.: Cosmological natural selection as the explanation for the complexity of the universe. Phys. A **340**, 705–713 (2004)
Steele, E.J.: Somatic Selection and Adaptive Evolution: On the Inheritance of Acquired Characters, 2nd edn. U. of Chicago Press, Chicago (1981)
Vaas, R.: Is there a Darwinian evolution of the cosmos? In: Proceedings of the MicroCosmos-MacroCosmos Conference, Aachen (1998)
Verhulst, J.: Discovering Evolutionary Principles through Comparative Morphology. Adonis Press, Ghent (2003)
Vermeij, G.J.: Historical contingency and the purported uniqueness of evolutionary innovations. PNAS **103**, 1804–1809 (2006)
Vermeij, G.J.: Nature: an economic history. Princeton U. Press, Princeton (2009)
Vidal, C.: Computational and biological analogies for understanding fine-tuned parameters in physics. Found. Sci. **15**(4), 375–393 (2010)
Wagman and Stephens: Surprising 'ultra-conserved regions discovered in human genome. UCSC Currents (2004)
West-Eberhard, M.J.: Developmental Plasticity and Evolution. Oxford U. Press, Oxford (2003)
Wilkins, A.S.: The Evolution of Developmental Pathways. Sinauer Associates, Sunderland (2001)
Yi, H., et al.: Gene expression atlas for human embryogenesis. FASEB J. **24**(9), 3341–3350 (2010)
Zimmer, C.: A Planet of Viruses, 2nd edn. U. Chicago Press, Chicago (2015)

Selecting Information in Financial Markets Herding and Opinion Swings in a Heterogeneous Mimetic Rational Agent-Based Model

Aymeric Vié[1,2(✉)]

[1] Sciences Po, Saint-Germain-en-Laye, France
aymericvie78@gmail.com
[2] University of Milan, Milan, Italy

Abstract. As expectations are driven by information, its selection is central in explaining common knowledge building and unraveling in financial markets. This paper addresses this information selection problem by proposing imitation as a key mechanism to explain opinion dynamics. Behavioral and cognitive approaches are combined to design mimetic rational agents able to infer and imitate each other's choices and strategies in opinion making process. Model simulations tend to reproduce stylized facts of financial markets such as opinion swings, innovation diffusion, social differentiation and existence of positive feedback loops. The influence of imitation reliability and information precision on opinion dynamics is discussed. The results shed light on two competing aspects of imitation behavior: building collective consensus and favoring innovation diffusion. The role of contrarian and individualistic attitudes in triggering large-scale changes is highlighted. From the results, some policy recommendations to reach better financial markets stability through opinion dynamics management are finally presented.

Keywords: Agent-based computational economics · Metamimetic chains
Mimetic rationality

1 Introduction

Motivation and Literature Review. Any dynamic economic system can be considered as an expectations feedback system, and any theory of expectation formation is a fundamental part of any economic model, notes [31]. The mechanisms behind beliefs formation in the context of financial markets has inspired novel approaches of heterogeneous agent models in finance using adaptive beliefs systems [7, 8, 17, 27]. More recently, Dosi et al. analyze heuristic rationality through the interconnexions between expectations and behaviors in evolving economies composed of heterogeneous agents [15]. These models reviewed by [19, 20] have achieved significant progress in reproducing real financial markets stylized facts, giving birth to endogenous bubbles and crashes corresponding to beliefs changes, and accounting for behavioral heterogeneity. The role of herding has been a constant focus as summarized by [1].

A. J. Morales et al. (Eds.): ICCS 2018, SPCOM, pp. 120–128, 2018.
https://doi.org/10.1007/978-3-319-96661-8_12

Some early herding mechanisms were introduced by [22, 29] to explain contagion and crowd effects in financial markets. Herding has recently been studied in the context of amplification and duration of financial crises [4] and increasing aggregate fluctuations in financial markets [12]. Herding seems adequate to explain evolutionary features of complex systems such as financial markets, but other behavior rules as payoff maximization and minority preference have been of great interest [3, 6, 9]. Indeed, due to time constraints and heuristics limitations, evolution relies on herding and more generally imitation rather than differential fitness as a primary evolutionary mechanism [16]. [13] provides a simple general definition of imitation mechanism: "A may want to be like B to a certain extent because from A's point of view, B is more successful". From these elements, we suggest imitation as a behavior mechanism of interest to investigate financial market dynamics. This approach is closely connected to Orléan's who combines rational and imitative behavior under "mimetic rationality". It develops models of mimetic contagion of information among investors in stock markets, relying on processes of opinion formation [29] and opening towards conventionalist rather than fundamentalist theories of asset pricing in financial markets, and autoreferential speculation mechanisms, in which the role of networks and neighbors is essential [18, 28, 31]. As we account for diversity of behavior by including these three decision rules in the model, our behavioral specification is closely related to the exploration of social interaction dynamics with heterogeneous agents that endogenously change their attitude to interact with the choices of others in [5]. This work is also closely related to the physics models developed by [25, 33]. Their model incorporates the endogenization of the sources of information onto the decisions of the agents, focusing on herding. This approach builds an Ising model combining herding, impact of external news and private information, where bounded rational agents following imitation mechanisms dynamically update their rules. Likewise, others include in their Ising model the possibility of local strategy changes driven by local information brought by nearest neighbors [21].

Purpose of the Research. We present a stochastic evolutionary game in which traders form expectations of the price where anticipations are driven by an information source choice, innovating by giving to our agents behavior heterogeneity and metamimetic cognition framework. A metamimetic chain [10] is defined as a chain of imitation metarules, which select some traits (strategies, characteristics) to be copied from the best agent according to a given criterion. The number of metarules involved is given by the cognitive bound. This mechanism satisfies three conditions: bounded rationality (metamimetic chains are finite according to a cognitive bound for each agent), metacognition (imitation rules become modifiable traits), and reflexivity (imitation rules can update reflexively changing the length of the metamimetic chain, or by updating themselves as the cognitive bound is reached). The high micro-level heterogeneity and locality of the interactions involved in the imitation mechanisms justify the use of agent-based modelling to study the dynamics of this metamimetic imitation system [2, 14, 26]. We address in this paper the problem of information selection between two constant information signals. Heterogeneous imitation processes with respect to neighborhood are implemented. The impact of expectations process performance over decision rules updates is analyzed. This paper is structured as follows. Section 2 presents the model formulation and main specifications. Section 3 presents our simulations results and discuss their interpretation and relation with previous research. Policy recommendations inspired from these results are also presented. Section 4 concludes.

2 Model Formulation

2.1 General Specifications

N traders are created and dispatched across a squared grid trading room. Time is divided into periods noted *t*, from 1 to T. At each period *t*, traders gather information through two different exclusive signal sources and observation of their close neighbors which is defined as a normalized radius around them. Traders then form daily expectations on the value of an asset relying on this information until period *T* that marks the end of the simulation.

2.2 Simulation Features

Traders Cognitive Abilities. The metamimetic chain [10] of each trader, i.e. its behavior rule set, is defined as $\{s_1, s_2, \ldots, s_k\}$ where s_k is the highest-order metarule. For any integer q, s_q takes one of the following entries we defined earlier: conformist, anticonformist and maximizer. The conformist (anticonformist) adopts the *inferred* most (least) used decision rule/information signal in his neighborhood defined as a radius. The maximizer trader adopts the *inferred* (i.e. what its infers to be) the decision rule/information signal used by the *inferred* most successful trader in sight. Inferring and perception mistakes are essential, as some noise is introduced from perception to action, throughout traders cognitive processes.

Information Signals. Two information signals φ (blue) and x (red) are defined. We denote φ_t and x_t the values of the respective signal at date t; φ_{it} and x_{it} the values of the public and private signals received by agent i at period t. Both signals are composed of a constant mean, and an idiosyncratic noise term noted ε. The blue and red noises are assumed to be independently and identically distributed among agents across time. $\varepsilon_{\varphi t}$ is has mean 0 and variance σ_{φ}^2, ε_{xit} has mean 0 and variance σ_x^2. Thus, at each period t and for each agent i, these signals take the following form:

$$\varphi_{it} = \omega + \varepsilon_{\varphi it} \tag{1}$$

$$x_{it} = \gamma + \varepsilon_{xit} \tag{2}$$

The initial choice of each agent between public or private information is determined according to a probabilistic process. The initial repartition of choices is exogenously set. We define θ_{it} is the weigh given by agent i to the private information at date t. For instance, if $\theta = 0.5$, the investor equally weighs the two sources of information, and will simply take the average of both to determine its expectation. If $\theta = 0$, the agent only considers the public information and will not rely on the private signal to form expectations. In our discrete case, θ is dummy and only takes the values 0 and 1. This setting allows each agent to rely exclusively on one of the two information sources.

2.3 Simulation Steps

At each date t, the following actions are executed: the price is determined according to the average of asset price expectations at the last date; the performance, i.e. the precision of the expectation with respect to price, of each agent is measured. This performance of neighbors is inferred and may trigger a metarule update from the observing agent. The information selection problem is solved by relying on the revised metarule. Eventually, receiving the information signals allow traders to form their expectation of the next price. So far, this model has assumed perfect vision and adaptiveness of “mind-readers” agents. More precisely, the agents are expected to correctly infer metarules, information choices and performance of their neighbors. Adding noise in actions and rules in the process may add realism to this idealistic hypothesis [10]. Such perceptive perfection may indeed not be naturally met in real financial markets. It is reasonable to expect some errors along the process: agents failing to identify a neighbor’s strategy or information choice or misestimating each other’s performance.

Price Determination. The price determination mechanism of the asset employed here is simple. It only aims to give investors a reference to estimate the performance of their information choice. The price at date $t+1$ is set as the average of traders’ last period anticipations of the price noted $\overline{E_t}(p_{t+1})$ to which we add a noisy idiosyncratic term $\varepsilon_{p_{t+1}}$ with mean 0 and variance $\sigma^2_{p_{t+1}}$.

$$p_{t+1} = \overline{E_t}(p_{t+1}) + \varepsilon_{p_{t+1}} \tag{3}$$

Performance Measuring. We define performance of agent i at period t noted π_{it} as the accuracy of agent i’s expectation formed at period $t-1$ with respect to the effective price p_t. Assuming that the performance function is normally distributed with a standard deviation of 20 around the mean p_t, adding a multiplication by 100 to allow a more visible performance differentiation between investors. Traders are allowed gains depending on their performance and are replaced by new traders if they run out of money, once a fixed cost (for being in the market) is introduced.

$$\pi_{it} = 100\frac{1}{20\sqrt{2\pi}}e^{-\frac{1}{2}\left(\frac{E_{i,t-1}(p_t)-p_t}{20}\right)} \tag{4}$$

The result of the performance evaluation function $\pi_{it}\left(E_{i,t-1}(p_t), p_t\right)$ is the “true” performance of agent i. The performance other agents perceive by observing an agent’s true performance is the inferred performance. This distinction becomes useful if agents can make mistakes in the observation process. Some noise can be added to neighbors’ performance inferring by agent i. The process by which agent j's performance is inferred by another agent i returns the exact performance value in the mistake-free situation. π_{jt} being a numeric value, both positive and negative errors can result from a mistaken perception. It can then be assumed than the result of the process is normally distributed with mean π_{jt} and variance σ_π^2 adding a gaussian term $\varepsilon_{\pi i}$ with mean 0 and variance σ_π^2. $\varepsilon_{\pi i}$ is idiosyncratic, independently and identically distributed among agents at each date. Noting $\widehat{\pi_{jt}}$ the inferred performance at period t of agent j by agent i:

$$\widehat{\pi_{jt}} = \pi_{jt} + \varepsilon_{\pi i} \tag{5}$$

Metarule Reflexive Update. At each period, traders test the adequacy of metarules with the state of the market. As we previously defined the set of metarules $\{s_1, s_2, \ldots, s_k\}$ where s_k is the highest-order metarule, this process can be formalized. As in our model the cognitive bound of the trader is equal to 1, its set of metarules only includes $\{s_k\}$. There is no endogenous variation of length of the metamimetic chain but only reflexive update of the unique decision rule: "rules for imitation can be their own metarule. The imitation rule reflexively updates by acting on itself as a modifiable trait." [10], testing its own adequacy with respect to itself. For instance, a Conformist investor would ask itself: "Is the Conformist metarule the most adapted from a Conformist point of view?". Neighbor's metarule inferring so far has been assumed to be a perfect process. Taking noise and mistakes into account, as behavior metarules are not numeric entries, adding a numeric noise would be irrelevant. Hence, it is more appropriate to introduce some error possibility observation process. We state that with a given probability ρ_O, the process returns a random different behavior among the two behaviors left.

Information Source Update. According to the lowest or unique metarule resulting from the previous step, the agents update their information source choice. The conformist agent chooses the more adopted information signal within its neighborhood. The anticonformist agent picks the least used information canal by its neighborhood. The maximizer agent relies on the information signal adopted by the more successful neighbor. As before, some noise is introduced in the inferring of the information source choice variable of agent j at a date t θ_{jt} by agent i. The process without noise always returns the right value of θ_{jt}. With a probability ρ_θ, the process returns a mistaken value of θ. As θ is dummy, the realization of the mistake will cause the process to return the wrong value of θ. The mistaken process will return 0 if $\theta_{jt} = 1$, and 1 when $\theta_{jt} = 0$ with a probability ρ_θ.

Agents' Price Expectations. In this final step, agents receive the information signal from the source they chose in the previous information source update step. They simply rely on this signal to form their expectation of the value of the price at the next date. We denote this naïve anticipation $E_{it}(p_{t+1})$.

$$E_{it}(p_{t+1}) = \theta_{it} x_{it} + (1 - \theta_{it}) \varphi_{it} \tag{6}$$

3 Results and Discussion

Stylized Facts Reproduction. Typical runs of the model implemented on Netlogo [32] give rise to stylized facts. Agents' successful forecasts reinforce the forecasts themselves, and where swings in opinions were endogenously generated [8, 23, 27, 30]. We observe wave-like innovation diffusion driven by positive feedback loops and percolation, endogenous clustering in groups of agents with similar opinions (information signals, in Fig. 1) and behaviors (Conformist in blue, Maximizer in red, Anticonformist in green) from random initial environments (Fig. 2).

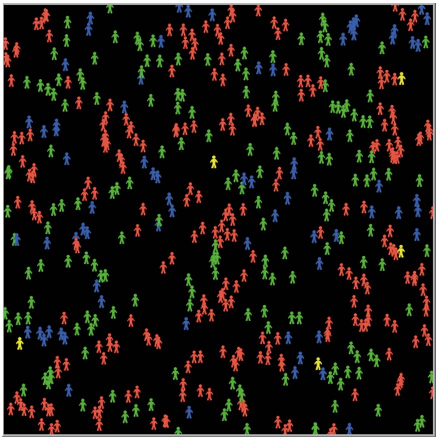


Fig. 1. Opinion clustering

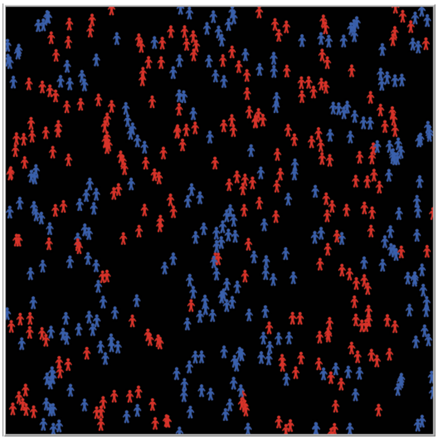


Fig. 2. Behavior clustering

Effects of Imitation Reliability and Information Precision. Despite the absence of complete information or traditional perfect sight and rationality, and accounting for behavior and cognitive heterogeneity, the market seems collectively efficient in adopting the least noisy signal, especially in the noise-free setting in imitation processes and converging to a collective adoption of the same information signal, that has the lowest idiosyncratic error term. As imitation reliability decreases by increasing inferring and imitation (opinion and behavior) mistake chances, the market struggles in settling to a counterfactually stable state. The information signal noise seems to only have a marginal impact on traders' performance given the high flexibility allowed by the metamimetic updates. However, a higher noise delta between signals increases the stability of the system (price volatility, information signals changes), although more strategical changes can be observed. Finally, the global imprecision parameter (composed of signals noise and price noise) has a very clear impact on the endogenous behavior distribution emerging from model runs and on market instability, as Table 1 below shows.

Table 1. Effect of global imprecision at T = 1000 over 100 runs (%) (From left to right: global imprecision, conformist/anticonformist/maximizer proportion, price changes, strategy changes, theta changes, information majority transition)

GI	Conf.	Anti.	Maxi.	P Ch.	St. Ch.	Θ Ch.	IT
0	93.367	1.70	4.93	0.13	65.14	37.40	2.6
1	91.62	2.00	6.37	2.35	77.80	45.09	7.45
2	88.06	2.49	9.44	4.60	98.27	56.72	9.55
3	81.21	3.51	15.28	6.81	136.49	79.19	9.75
4	74.57	4.41	21.02	9.03	169.75	98.89	16.1
5	61.02	6.29	32.69	11.43	239.85	142.39	26.65
6	52.05	7.46	40.48	13.69	282.55	168.77	29.3
7	47.81	7.94	44.24	15.87	299.80	180.40	29.3
8	52.03	7.26	40.71	18.43	273.99	163.76	26.25
9	46.95	7.99	45.05	20.44	298.53	178.87	30.2

4 Conclusion

We introduce an agent-based model for beliefs formation in financial markets with information selection and performance differentiation. Agents with bounded rationality and limited sight are collectively efficient in choosing the least noisy information signal only through imitation. Our results shed light on two competing aspects of imitation in building efficient common knowledge and favoring innovation diffusion. We qualify the impact of herding behavior by outlining the initial contribution of payoff-based imitators triggering opinion dynamics. Our Monte Carlo simulations show that through imitation, investors are quite efficient in adopting the most precise, or at least the least noisy information signal. This highlights the need for a very precise and unequivocal "public information", and specifically applies for public financial regulation authorities. Future research should investigate the possibilities opened by the application of Metamimetic Games to these systems, comparing to traditional models, and study more in detail model results through Monte Carlo simulation and sensitivity analysis.

References

1. Alfarano, S., Lux, T., Wagner, F.: Estimation of agent-based models: the case of an asymmetric herding model. Comput. Econ. **26**(1), 19–49 (2005)
2. Arthur, B.W.: Complexity and the economy. Oxford University Press, Oxford (2014)
3. Arthur, B.W., Holland, J.H., Lebaron, B., Palmer, R.G., Tayler, P.: Asset Pricing under Endogenous Expectations in an Artificial Stock Market (1996)
4. Assenza, T., Brock, W.A., Hommes, C.H.: Animal spirits, heterogeneous expectations, and the amplification and duration of crises. Econ. Inq. **55**(1), 542–564 (2017)
5. Barucci, E., Tolotti, M.: The dynamics of social interaction with agents' heterogeneity (2009)
6. Brock, W.A., Durlauf, S.N.: Discrete choice with social interactions. Rev. Econ. Stud. **68**(2), 235–260 (2001)

7. Brock, W.A., Hommes, C.H.: Heterogeneous beliefs and routes to chaos in a simple asset pricing model. J. Econ. Dyn. Control **22**(8–9), 1235–1274 (1998)
8. Brock, W.A., Hommes, C.H.: Rational animal spirits. In: Herings, P.J.J., van der Laan, G., Talman, A.J.J. (eds.) The Theory of Markets, North-Holland, Amsterdam, pp. 109–137 (1999)
9. Challet, D., Marsili, M., Zhang, Y.C.: Minority games: interacting agents in financial markets. OUP Catalogue (2013)
10. Chavalarias, D.: Metamimetic games: modeling metadynamics in social cognition. J. Artif. Soc. Soc. Simul. **9**(2), 5 (2006). http://jasss.soc.surrey.ac.uk/9/2/5.html
11. Chiarella, C., He, X.-Z.: Heterogeneous beliefs, risk and learning in a simple asset pricing model. Comput. Econ. **19**(1), 95–132 (2002)
12. Cont, R., Bouchaud, J.P.: Herd behavior and aggregate fluctuations in financial markets. Macroecon. Dyn. **4**(2), 170–196 (2000)
13. Conte, R., Paolucci, M.: Intelligent social learning. J. Artif. Soc. Soc. Simul. **4**(1), U61–U82 (2001). http://www.soc.surrey.ac.uk/JASSS/4/1/3.html
14. Daudé, E.: Contributions of multi-agent systems for diffusion processes studies. Cybergeo: Eur. J. Geogr. **255**, 1–16 (2004)
15. Dosi, G., Napoletano, M., Roventini, A., Stiglitz, J., Treibich, T.: Rational Heuristics? Expectations and behaviors in Evolving Economies with Heterogeneous interacting agents (2017)
16. Frank, H.: Natural selection, rational economic behavior and alternative outcomes of the evolutionary process. J. Socio-Econ. **32–6**(12), 601–622 (2003)
17. Gaunersdorfer, A.: Endogenous fluctuations in a simple asset pricing model with heterogeneous agents. J. Econ. Dyn. Control **24**, 799–831 (2000)
18. Harras, G., Sornette, D.: How to grow a bubble: a model of myopic adapting agents. J. Econ. Behav. Organ. **80**(1), 137–152 (2011)
19. Hommes, C.H.: Financial markets as nonlinear adaptive evolutionary systems (2001)
20. Hommes, C.H.: Heterogeneous agent models in economics and finance. In: Tesfatsion, L., Judd, K.L. (eds.) Handbook of Computational Economics, vol. 2, pp. 1109–1186. Elsevier (2006). Chap. 23
21. Kaizoji, T., Bornholdt, S., Fujiwara, Y.: Dynamics of price and trading volume in a spin model of stock markets with heterogeneous agents. Phys. A: Stat. Mech. Appl. **316**(1), 441–452 (2002)
22. Kirman, A.: Ants, rationality, and recruitment. Q. J. Econ. **108**(1), 137–156 (1993)
23. Kirman, A.P., Teyssiere, G.: Micro-economic models for long memory in the volatility of financial time series, In: Herings, P.J.J., Van der Laan, G., Talman, A.J.J. (eds.) The Theory of Markets, North Holland, Amsterdam, pp. 109–137 (2002)
24. Kirman, A., Zimmermann, J.B. (eds.) Economics with Heterogeneous Interacting Agents, vol. 503. Springer Science and Business Media (2012)
25. Kristoufek, L., Vosvrda, M.: Herding, minority game, market clearing and efficient markets in a simple spin model framework. Commun. Nonlinear Sci. Numer. Simul. **54**, 148–155 (2017)
26. Lux, T.: Stochastic behavioral asset-pricing models and the stylized facts (2009)
27. Lux, T., Marchesi, M.: Scaling and criticality in a stochastic multiagent model of a financial market. Nature **397**, 498–500 (1999)
28. Makarewicz, T.: Contrarian behavior, information networks and heterogeneous expectations in an asset pricing model. Comput. Econ. **50**(2), 231–279 (2017)
29. Orléan, A.: Bayesian interactions and collective dynamics of opinion - herd behavior and mimetic contagion. J. Econ. Behav. Organ. **28**, 257–274 (1995)

30. Sornette, D., Zhou, W.X.: Importance of positive feedbacks and overconfidence in a self-fulfilling Ising model of financial markets. Phys. A: Stat. Mech. Appl. **370**(2), 704–726 (2006)
31. Tsakas, N.: Naive learning in social networks: Imitating the most successful neighbor (2012)
32. Wilensky, U.: NetLogo (1999)
33. Zhou, W.X., Sornette, D.: Self-organizing Ising model of financial markets. Eur. Phys. J. B **55**(2), 175–181 (2007)

Election Methods and Collective Decisions

Thomas Edward Cavin(✉)

Department of Social Science and Policy Studies, Salisbury Laboratories, Worcester Polytechnic Institute, 100 Institute Road, Worcester, MA 01609-2280, USA
tec@wpi.edu

Abstract. This paper presents some simulation results on various collective decision methods in the context of Downsian proximity electorates. I show why these results are less than ideal, and contrast these different voting systems with a new system called Serial Approval Vote Elections (SAVE), which produces better outcomes that approach the ideal represented by the median voter theorem. I show how SAVE works in both normal and unusual electorates, how SAVE can be easily integrated into committee procedures, and how SAVE can be used in larger elections.

Keywords: Collective decisions · Voting methods
Median voter theorem · Arrow's impossibility theorem
Serial approval vote elections

1 Introduction

The purpose of this paper is to briefly introduce a new collective decision making system that is designed to find the will of the people. In essence, this paper can be considered the technical specifications of this system as a sensor, in terms of how accurate its readings are, and what is needed to get those readings.

2 Representation of the Electorate

To analyze collective decisions in simulation, I needed a way to represent what the people want. Historically voting method papers have used preference profiles to represent what the voters want (Dodgson [7], Black [2], Arrow [1], etc.). Table 1 shows an example of a preference profile depicting a situation in which there is no best answer for the location of a new communial resource. In any pairwise comparison, one of the alternatives is clearly preferred over the other. Yet in each case the excluded alternative is strongly preferred over the winner of the pairwise comparison. This is a classic majority cycle, and such cycles are the root issue underlying Arrow's impossibility theorem [1]. The profile tells us this

A. J. Morales et al. (Eds.): ICCS 2018, SPCOM, pp. 129–137, 2018.
https://doi.org/10.1007/978-3-319-96661-8_13

Table 1. A basic example of a preference profile showing a majority cycle.

Region	Preferences	Voters	Region	Preferences	Voters	Region	Preferences	Voters
1	$A \succ B \succ C$	25	3	$B \succ A \succ C$	5	5	$C \succ A \succ B$	25
2	$A \succ C \succ B$	10	4	$B \succ C \succ A$	30	6	$C \succ B \succ A$	5

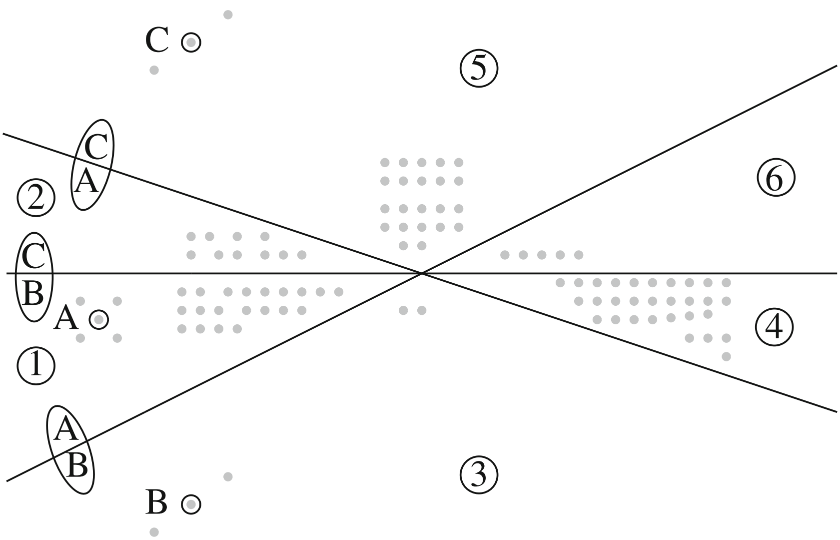


Fig. 1. One of many possible Downsian proximity electorates showing a distribution of voter optimums that would generate the preference profile shown in Table 1.

situation is bad, but it provides no clue as to why it is bad. For that we need more data.

In contrast, Fig. 1 illustrates one possible electorate configuration that generates the voter preference profile in Table 1. In this figure each gray dot marks one voter's ideal location of a new communal resource. Three of these locations have been proposed as indicated by circles around the locations labeled A, B, and C. All voters prefer a location closer to their personal ideal over a location further away. The three lines divide the issue space into six regions, labeled to match the voter preferences in Table 1. Any rearrangement of the voter ideal positions that do not cross any lines will generate this same preference profile as will many other Downsian proximity electorates.

This Downsian electorate figure presents much more data than the preference profile table, and also provides more insight into voter behavior. While the preference profile brings attention to Condorcet winners, losers and cycles, the Downsian representation highlights the center, the edges, and bands that are roughly the same distance from the center. This visualization provides intuitive support for the median voter theorem, credited to both Black [2] and Downs [8]. The basic concept is an alternative at the center of the ideal positions in a

Downsian electorate will defeat an alternative distant from that center, simply because any significant deviation from the center in any direction will only be favored by a minority and thus will be opposed by a majority.

3 Standard Voting Methods

I chose the following voting methods for comparison because they are either commonly used (simple majority, honest and strategic plurality) or being promoted as solution to the problems of plurality voting (IRV).

Simple Majority (SM): Voters vote for only one of two alternatives, the ballots are tallied, and the alternative with a majority of the votes is the winner.

Honest Plurality (HP): Voters vote for only one of multiple alternatives, the ballots are tallied, and the alternative with the most votes is the winner.

Strategic Plurality (SP): Same rules as HP, but voters have an expectation of the final result and may vote strategically to try to get a better result.

Instant Runoff Voting (IRV): This method is described in Feld [10] and Cary [5]. Basically, voters submit a ballot indicating a preference order of the alternatives. The ballots are counted in a series of rounds, and the alternative with the smallest vote count is eliminated and its ballots redistributed until one alternative receives a majority.

Best Initial Alternative: This is an artificial proxy method for all other possible single-round voting methods. It simply selects the initial alternative with the best centrality score (see Sect. 5.1) as the winner.

All of the above voting procedures are limited to the initial alternatives and do not specify how alternatives get on the ballot. The next, new voting method does not share these limitations.

4 Serial Approval Vote Elections

Serial Approval Vote Elections (SAVE) is detailed in Cavin and Pavlov [6], but for the purposes of this paper, it is sufficient to know that the procedure is iterative and allows the introduction of new alternatives during the process. In essence, SAVE starts with multiple initial alternatives, selects one of them to be the focus, and then iterates to improve the focus. At each iteration voters indicate which alternatives are better than the focus, whether the focus is good enough to keep, and have the option to introduce new alternatives. SAVE is done when more voters want the current focus as the winner than want to change the focus. The final winner may be a Condorcet winner, part of a majority cycle, or simply good enough.

Once the final winner has been determined, an optional but strongly suggested final mandate round using approval voting [3] is held that explicitly and independently measures the support of the winner and all other alternatives.

5 Test Environment - Electorate Generation

I use synthetic Downsian electorates for my tests. Each basic electorate is determined by its number of voters, number of issues, and a random seed. A single voter has two values for each issue: an opinion normally distributed around zero, and a weight between zero and the number of issues. For each voter the sum of its issue weights equals the number of issues. I also create what I term lumpy electorates using two additional parameters: the number of subgroups and the minimum size of each subgroup. For these electorates, the minimum number of voters are allocated to each subgroup, then the number of remaining voters is randomly partitioned using an algorithm from Nijenhuis and Wilf [13]. The subgroup allocations are filled by first creating one voter for each subgroup using the normal distribution. This first set of voters are referred to as parents. Then the remaining voters for each subgroup are created by generating voters normally and offsetting each voter's values by the corresponding value from its subgroup parent. Thus each subgroup is normally distributed, but has a different mean from the electorate as a whole.

I worked with lumpy distributions because they are more complicated than normal distributions, and have more structure than uniform distributions. They also have the added benefit of tending to be less dense in the center. I considered this to be beneficial because my task was to develop a process to find the center and I did not want to make it easy.

5.1 Centrality Score

The center of an electorate is not well-defined, or rather there are many well-defined centers that do not always coincide. I use a special cluster of three different centers for my metric: the mean position, the median position, and the position I call the gradient descent or gradient. The mean position is the standard weighted mean from statistics. The median position is the weighted median as defined in Gurwitz [11]. More specifically, this is the dimension-by-dimension median that was proven to be a majority rule equilibrium point under certain conditions in Humphreys and Laver [12]. The gradient descent position is numerically determined using a hill climbing algorithm that starts at an arbitrary position P in the issue space, calculates the vector sum of the inward pointing unit vectors normal to the indifference curve through P for each member, moves P a bit in that direction, and iterates until the vector sum is essentially zero.

The mean edge length of the special cluster D_S is a type of a diameter, and that distance is used with the electorate diameter D_E of six standard deviations to create a crude measure of how well-defined the center of the electorate is, using $(D_E - D_S)/D_E$ expressed as a percentage. The centrality score of an alternative C_a is similarly defined by the mean radius R_a, or the mean distance $d(a, S)$ from a to the special cluster positions, corrected for the fact that the special cluster is a diameter and spans the center: $R_a = d(a, S) - D_S/2$. The centrality score is then $C_a = 1 - R_a/R_E$ expressed as a percentage and where $R_E = D_E/2$ or three standard deviations of the electorate.

6 Results

My test results are given in the following tables. Where pairs are given in a cell separated by a slash they are *mean/stdev*. Unless otherwise stated, the initial alternatives in any elections are the subgroup parents for lumpy electorates. Any unusual table headings are explained in the text.

Table 2. Comparison of basic voting methods.

Voting method	Vote count	Vote %	Mean radius	Centrality %
Honest plurality	89.0/21.7	17.8/4.3	0.93/0.36	56.4/16.2
Simple majority A	295.9/30.7	59.2/6.1	0.58/0.28	73.1/11.6
Simple majority B	293.0/32.0	58.6/6.4	0.71/0.31	66.5/13.2
Strategic plurality A	276.7/32.9	55.3/6.6	0.58/0.28	72.9/11.9
Strategic plurality B	276.7/32.7	55.3/6.5	0.71/0.30	66.4/13.2
Instant runoff	284.1/30.3	56.8/6.1	0.55/0.23	73.8/10.5
Best initial	−/−	−/−	0.33/0.15	84.4/6.9
SAVE	318.0/52.2	63.6/10.4	0.09/0.08	95.6/3.4

In Table 2, the $A|B$ markers indicate the method used to determine the two best alternatives: A is based on subgroup size and B is based on the honest plurality vote. I attribute the significantly worse centrality of the B rows to the poor centrality of the honest plurality winners. Using the subgroup parents as initial alternatives is not intrinsically bad as shown by the best initial centrality scores, but none of the standard voting methods are very good at selecting the best one. The SAVE results, however, are clearly and significantly better than even the best initial and are very close to the center. However, the SAVE results come at a cost, as we see in the next table.

Table 3. The effects of number of issues on SAVE results.

Issues	Rounds	Specials %	Best initial %	SAVE %	Gain %
2	17.0/7.3	98.4/0.8	89.8/5.7	97.6/3.0	+7.8/5.9
3	20.9/8.6	97.7/0.9	83.0/6.8	94.8/4.7	+11.8/7.3
4	23.5/8.9	96.9/1.1	76.9/7.2	92.0/5.5	+15.1/8.1
5	25.1/9.6	95.9/1.4	71.5/8.0	88.9/6.7	+17.4/9.0
6	26.1/9.6	95.0/1.6	66.8/8.7	86.1/7.5	+19.2/10.0
all	22.5/9.4	96.8/1.7	77.6/11.0	91.9/7.0	+14.2/9.2

In Table 3 we see how SAVE responds to increasing the number of issues in the electorate. Basically, as the number of issues increase, the centrality of the special cluster decreases indicating the center is less well-defined, and the round count increases slightly. The best initials column indicates the parents are less likely to be central as the issues increase, and the SAVE centrality percentage also shows this decrease. However as the gain percentage shows, the fractional gain of using SAVE increases with the increased complexity of multiple issues. The next table looks at the number of voters.

Table 4. The effect of electorate size on SAVE performance.

Voters	Rounds	Specials %	Best initial %	SAVE %	Gain %
200	15.8/3.8	95.6/0.9	77.1/6.4	88.9/5.4	+11.7/7.7
400	18.9/4.9	96.7/1.2	68.4/10.3	90.4/3.1	+22.0/11.1
800	19.8/7.5	97.6/0.9	71.4/9.6	90.4/2.3	+19.0/10.1
1600	20.8/4.7	97.5/0.9	72.0/5.1	90.2/5.1	+18.2/6.8
3200	22.9/6.1	97.1/2.5	69.1/12.4	92.2/4.8	+23.0/12.0
6400	25.5/15.1	97.7/0.9	72.4/10.7	92.9/2.3	+20.5/10.7

In Table 4 we see how the performance of SAVE varies with the size of the electorate. SAVE is almost indifferent to changes in electorate size, which is what I expected from a parallel algorithm and what I would require from a voting system.

Table 5. Effects of quality of initial alternatives on SAVE performance.

Issues	Rounds	Specials %	Best initial %	SAVE %	Gain %
2	17.0/7.3	98.4/0.8	89.8/5.7	97.6/3.0	+7.8/5.9
2 rerun	13.2/6.0	98.4/0.8	99.0/1.3	98.6/1.5	−0.4/1.1
3	20.9/8.6	97.7/0.9	83.0/6.8	94.8/4.7	+11.8/7.3
3 rerun	15.3/7.3	97.7/0.9	97.2/2.6	96.7/2.7	−0.5/1.5
4	23.5/8.9	96.9/1.1	76.9/7.2	92.0/5.5	+15.1/8.1
4 rerun	16.7/8.3	96.9/1.1	95.0/3.4	94.5/3.6	−0.5/1.8
5	25.1/9.6	95.9/1.4	71.5/8.0	88.9/6.7	+17.4/9.0
5 rerun	17.8/9.2	95.9/1.4	92.6/4.4	92.1/4.4	−0.5/2.3
6	26.1/9.6	95.0/1.6	66.8/8.7	86.1/7.5	+19.2/10.0
6 rerun	18.5/9.9	95.0/1.6	90.2/4.8	89.8/5.0	−0.4/2.4
all	22.5/9.4	96.8/1.7	77.6/11.0	91.9/7.0	+14.2/9.2
all rerun	16.3/8.4	96.8/1.7	94.8/4.7	94.3/4.8	−0.5/1.9

In Table 5 we see what happens when SAVE is run a second time in an electorate. The rows without "rerun" are taken from Table 3, with the new data indicated by the label "rerun". The only difference between the runs is that in addition to the subgroup parents as initial alternatives, the competitive alternatives from the end of the first run are added as initials in the rerun. In this experiment the round count drops significantly, the best initials are extremely good, the SAVE centrality is better, and the gain over the best initials is a slight loss on average. I attribute this loss mainly to the fact that the best initial alternative is really good, and further note the SAVE result is still generally better than it was the first time and was found in far fewer rounds.

7 Comments

I've shown SAVE works well and gets significantly better results than existing systems, though at the cost of more time and effort. SAVE does well as trade-offs among multiple issues become more relevant. SAVE shows significant performance gains after its first use, which suggests it may be easier to track the center than initially find it. Further, SAVE shows no discernible trends as the size of the electorate increases, which indicates it scales well and should be usable over a wide range of sizes.

At the smaller end of the decision making spectrum, the committee procedures described by Black [2], are already iterative, so a switch to SAVE could be done with very little disruption to procedures. The only major effects would come from outcomes that better reflect the opinions of the full committee.

On a larger scale, such as the US presidential election, SAVE could replace the primaries, conventions, and the final election by starting a year or two before the next term would begin, and having a round once every two to four weeks. No candidates would ever be eliminated, and the focus would only improve from wherever it started. I presume both voters and candidates would learn during the process. Voters would have time to become better informed about the candidates, and the candidates would get fairly direct feedback on what the voters collectively want.

One valid question regarding SAVE is whether large groups of people will embrace a collective decision process that takes so many rounds. As there have been no field trials of SAVE so far, I have no data to answer this question. However, I feel optimistic for the following reasons. Social scientists in public choice have looked at the question of why people vote or what drives voter turnout, and have two rival accounts of rational voting: instrumental theory and expressive theory [4,9]. Roughly, the instrumental theory is that voters choose to participate when their vote matters, and the expressive theory is that voters choose to vote to support their team. Both these theories would support turnout in SAVE because expressively motivated voters have many opportunities to support their issues or teams, and instrumentally motivated voters are making a difference with their votes because SAVE vote counts (particularly in the later rounds) tend to be very close so every vote matters. Both theories of voter turnout assume voters behave rationally, choosing to vote when benefits outweigh costs and to not vote when costs outweigh benefits. The multiple rounds of voting in SAVE provide opportunities to reduce many of the costs of voting, which should increase turnout regardless of the reasons for participating.

A second valid question regarding scaling is the mechanics of the nomination process. In a committee with every member in the same room, proposing a new alternative that combines existing alternatives is reasonably easy. As the group gets larger, the same procedure can be used but the logistics of informing the electorate of the new alternatives gets a bit more complicated. A free and open internet can facilitate this process, but the exact mechanisms need to be specified. Further, the only examples we currently have for really large elections are our elections for individuals to hold public offices. Under plurality rules such elections end up conflating the individuals running for the offices with the policy goals of those offices. I would prefer we use SAVE to set our policy goals first. Then once our policy goals are set, we use SAVE again to elect the individuals to administer those policies.

I think are other potential consequences of using SAVE that may be even more important over the long run than getting better policies and administrators. Our current plurality voting rules encourage strategic voting because in that system being strategic gets better results than being strictly honest. Under SAVE, an honest vote is a strategic vote, so there is real motivation to be truthful. The ease of introducing new alternatives makes it easier to address emergent issues, which may make it more likely that we could start work to correct problems while they are still small and easier to resolve. Plurality voting rules structurally encourage division and antagonism due to their zero-sum nature. Under these rules different special interests need to form exclusive alliances in order to have any chance of getting their needs or desires addressed. Given the instability of the plurality voting system with more than two parties, smaller or less powerful interest groups are structurally limited in those alliances to one of two opposing political groups. This is a recipe for conflict. In contrast SAVE operates under a paradigm of structural cooperation. Instead of two large opposing groups in perpetual conflict, SAVE automatically coordinates vast numbers of small groups and individuals to reach a collective balance of the needs and desires of all the different and varied groups in the electorate. The social possibilities seem almost beyond limit.

Yet the possibilities of SAVE are just that, possibilities, until we collectively make them real. We are currently faced with many complex problems. Some of those problems are of our own making, while others are not. But all of our problems will be easier to overcome if we can work together to both find and implement their solutions. SAVE will not solve our problems. But it may be the tool we need to implement our solutions.

References

1. Arrow, K.J.: Social Choice and Individual Values, Cowles Foundation Monograph, 3rd edn., vol. 12. Yale University Press (2012)
2. Black, D.: On the rationale of group decision-making. J. Polit. Econ. **56**(1), 23–34 (1948). http://www.jstor.org/stable/1825026

3. Brams, S.J., Fishburn, P.C.: Approval voting. Am. Polit. Sci. Rev. **72**(3), 831–847 (1978). https://doi.org/10.2307/1955105. http://www.jstor.org/stable/1955105
4. Brennan, G., Hamlin, A.: Expressive voting and electoral equilibrium. Public Choice **95**(1–2), 149–175 (1998). http://link.springer.com/article/10.1023/A:1004936203144
5. Cary, D.: Estimating the margin of victory for instant-runoff voting. In: Electronic Voting Technology Workshop/Workshop on Trustworthy Elections, San Francisco, CA, (2011). https://www.usenix.org/events/evtwote11/tech/final_files/Cary8-2-11.pdf
6. Cavin, T.E., Pavlov, O.V.: Social Decisions in a Democracy and the Introduction of Serial Approval Vote Elections. Social Choice and Welfare (2018, submitted)
7. Dodgson, C.L.: A Method For Taking Votes on More than Two Issues. Clarendon, Oxford (1876). Reprint in Black, D.: The Theory of Committees and Elections, pp. 224–234. Cambridge University Press, Cambridge, Mass. (1958) and McLean, I., Urken, A.: Classics of Social Choice. University of Michigan Press, Ann Arbor, Michigan (1995)
8. Downs, A.: An Economic Theory of Democracy. Harper, New York (1957)
9. Feddersen, T.J.: Rational choice theory and the paradox of not voting. J. Econo. Perspect. **18**(1), 99–112 (2004). http://www.jstor.org/stable/3216877
10. Grofman, B., Feld, S.L.: If you like the alternative vote (aka the instant runoff), then you ought to know about the Coombs rule. Elect. Stud. **23**(4), 641–659 (2004). http://www.sciencedirect.com/science/article/pii/S026137940300060X
11. Gurwitz, C.: Weighted median algorithms for L_1 approximation. BIT **30**(2), 301–310 (1990). https://doi.org/10.1007/BF02017350. https://link.springer.com/article/10.1007/BF02017350
12. Humphreys, M., Laver, M.: Spatial models, cognitive metrics, and majority rule equilibria. Br. J. Polit. Sci. **40**(01), 11–30 (2010). http://journals.cambridge.org/abstract_S0007123409990263
13. Nijenhuis, A., Wilf, H.S.: Combinatorial algorithms (1975). http://cds.cern.ch/record/104492

Agent-Based Models for Assessing Social Influence Strategies

Zachary K. Stine(✉) and Nitin Agarwal

University of Arkansas at Little Rock, Little Rock, AR 72204, USA
{zkstine,nxagarwal}@ualr.com

Abstract. Motivated by the increasing attention given to automated information campaigns and their potential to influence information ecosystems online, we argue that agent-based models of opinion dynamics provide a useful environment for understanding and assessing social influence strategies. This approach allows us to build theory about the efficacy of various influence strategies, forces us to be precise and rigorous about our assumptions surrounding such strategies, and highlights potential gaps in existing models. We present a case study illustrating these points in which we adapt a strategy, namely, amplification, commonly employed by so-called 'bots' within social media. We treat it as a simple agent strategy situated within three models of opinion dynamics using three different mechanisms of social influence. We present early findings from this work suggesting that a simple amplification strategy is only successful in cases where it is assumed that any given agent is capable of being influenced by almost any other agent, and is likewise unsuccessful in cases that assume agents have more restrictive criteria for who may influence them. The outcomes of this case study suggest ways in which the amplification strategy can be made more robust, and thus more relevant for extrapolating to real-world strategies. We discuss how this methodology might be applied to more sophisticated strategies and the broader benefits of this approach as a complement to empirical methods.

Keywords: Social influence · Opinion dynamics
Automated information campaigns · Social bots

1 Introduction

1.1 Bots and Automated Information Campaigns

Concerns about the ability to conduct large-scale opinion manipulation through social media have existed for several years now, but continue to grow [1–3]. Many of these concerns highlight the use of automated social media accounts to artificially amplify some message, the assumption being that more people will adopt the viewpoint espoused by the message than would hold that opinion otherwise. These automated accounts are variously referred to as "bots," "social bots," and "influence bots," among others.

Many studies exist which describe how bots propagate various messages, perhaps in the form of links to articles, but the effects these bots have on a population's opinions is

A. J. Morales et al. (Eds.): ICCS 2018, SPCOM, pp. 138–146, 2018.
https://doi.org/10.1007/978-3-319-96661-8_14

poorly understood, owing to the difficulty of measuring such an effect. While studies exist which show that misinformation is able to spread quickly within social media while also noting the presence of bots attempting to propagate misinformation [4], it is difficult to say what degree of influence bots have on the social spaces in which they operate. In the absence of useful empirical work about the effects of such a strategy, we posit that theoretical knowledge gained from agent-based models is useful, despite the simplifications involved in moving the focus from a human population to artificial agents.

1.2 Agent-Based Models of Social Influence

While there is tremendous variety among existing models of social influence, they all generally describe a process by which a population of agents, each possessing some state, change their states over time as a result of interacting with other agents. Agent states are typically likened to opinions and attitudes; thus, social influence models are analogized to the changing of opinions within a population over time. Three broad classes of social influence models are identified in [5]: assimilative influence, similarity-biased influence, and repulsive influence.

Assimilative Influence. Agents that behave according to assimilative influence always reduce their differences in interactions, typically resulting in the entire population converging at a consensus state.

Similarity-Biased Influence. Similarity-biased influence models add some requirement to this process, typically in the form of a distance threshold: if agent *i*'s state is within some distance from agent *j*'s state, then *j*'s state will update to become closer (i.e. more similar) to that of agent *i*; otherwise, agent *j*'s state remains unchanged. This type of model is also referred to as the Deffuant model [6], and is also referred to as a bounded confidence model, following [7]. Depending on the similarity threshold used, models of this type may result in agents reaching a consensus state just as in assimilative influence, but as the criteria for influence becomes stricter, agents may become divided into clusters that remain stable over time.

Repulsive Influence. Repulsive influence models go a step further, allowing agents to become more dissimilar depending on some defined criteria. Repulsive influence models are capable of generating bipolarization, in which the agents split themselves into two clusters on the extreme poles of the range of possible states. Of the three classes, repulsive influence is the only one that can result in agents moving outside their initial range of states [5]. We assess the same social influence strategies using each of the three mechanisms and compare the results.

1.3 Opinion Control as Automated Information Campaign

A highly analytical study of how the opinion dynamics produced in similarity-biased models (or bounded confidence models) is given in [8]. In this study, we carry out similar work, but extend the behaviors analyzed to also include assimilative influence and repulsive influence. The case of repulsive influence will prove to be particularly interesting in that it is capable of generating behavior which may seem counterintuitive.

We do not contend that there is a clear, one-to-one relationship between conclusions drawn from the models we describe to the actual automated information campaigns and their effects on people, which motivate this study. However, agent-based models are fully capable of describing any bot observed in the real world. This is true since a bot is an automated account; thus, its behavior must be describable by a computer program. The same is true of the artificial agents which populate these models. The relationship between humans and artificial agents is much less clear. There is a great deal that can be learned about automated information campaigns from agent-based models, though it requires care in interpreting results. Many excellent studies exist which explore the ways in which agent-based models are useful despite their potentially complicated relationship to the "real-world," and we suggest interested readers consult [9, 10].

2 Model

2.1 Neutral Population

Each of the experiments we describe in this paper are simulated in exactly the same way, except for the particular social influence mechanism used. We use the discrete set of possible agent states, $\{-1, -0.99, \ldots, 0.99, 1\}$. In each model run, there is a neutral population of 100 agents. These agents are neutral in that they are not engaging in any opinion manipulation strategy—they simply share their own opinions with complete transparency and update their opinions according to the particular social influence mechanism being used. Each agent is initialized with an opinion that is uniformly distributed across the opinion range. Each run consists of 5,000 time steps. At each time step, two random orderings of agents are generated. One specifies the order of agents to play the role of sender, the other specifies the order of agents to be receivers (with the constraint that the same agent cannot be in the same spot in line in both lists). In the random order just specified, each sender and receiver are paired to interact.

In the case of assimilative influence, the receiver's opinion is updated to be the next possible state closer to the sender (*i.e.* the new opinion will either be 0.01 more or less than the current opinion, depending on which brings the receiver closer to the sender). In the case of similarity-biased influence, if the distance between the sender's opinion and the receiver's opinion is greater than the threshold parameter, ε, then the receiver does not update its opinion. If the distance between their opinions is less than ε, the receiver updates its opinion to be closer to the sender. However, if the distance is equal to ε, either option—update to be closer or stay the same—is randomly chosen with equal probability. The case of repulsive influence is identical to that of similarity-biased influence, except that the receiver's state will be updated to the next possible state *further away* from the sender, if either the distance between the sender and receiver's opinions is greater than ε, or if the distance is equal to ε in which case the receiver's state will become either closer or further away from that of the sender with equal probability. In all cases, if the receiver's state is the same as the sender's state, then the receiver's state is not updated.

2.2 Manipulative Agents and Influence Strategies

In order to assess the efficacy of an influence strategy, such a strategy must be sufficiently well-specified so that it can be implemented as agent behavior. In this paper, we are particularly interested in assessing amplification strategies, motivated by the observed behavior of bots being used to amplify certain messages. To do this, we add some number of manipulative agents to the neutral population. These manipulative agents are not subject to the same rules as the neutral agents. Instead of both sending and receiving opinions, they only send opinions; the rationale being that they are bot-like agents, incapable of having their opinions changed through interactions. Instead, they only exist to amplify some specified opinion. We assume that neutral agents are incapable of distinguishing between neutral and manipulative agents.

In simulation runs with manipulative agents present, the same steps take place as described for the neutral population, except that, after all neutral sender/receiver pairs have interacted in a time step, each manipulative agent is randomly paired with a neutral agent that receives the manipulative agent's opinion. Thus, each neutral agent is still able to act as a sender with potential influence once each time step, but some neutral agents will play the role of receiver multiple times: once with a neutral sender and one or more times with a manipulative sender. There is no constraint that the neutral agents being manipulated must be unique each time step, so the possibility exists that they may be randomly chosen multiple times in a single time step.

In each strategy (described below), we define an amplification parameter, α, to denote the number of manipulative agents as a proportion of the neutral population. For example, if α is 0.05, then five manipulative agents are used in the simulation run. Another way to interpret α is as an approximation of the proportion of neutral agents that will be potentially manipulated within each time step. Thus, an increase in a strategy's amplification, α, corresponds to an increase in the proportion of neutral agents which are manipulated in each time step.

To assess a strategy's efficacy, we follow [8] in defining a target opinion, though we limit this target opinion to a single state: 1. This allows us to score a strategy based on the proportion of neutral agents that possess the target opinion at the end of a simulation run. Thus, a score of 1 indicates that the entire neutral population held the target opinion at the end of a simulation run. We assess each strategy within the context of assimilative influence, similarity-biased influence, and repulsive influence. For the cases of similarity-biased and repulsive influence, we examine results using two values of the threshold parameter, ε: 0.9 and 0.6. For each unique set of model parameters, we run 500 simulations, from which average scores are calculated.

Simple Amplification Strategy. To assess a simple amplification strategy, the manipulative agents simply send the target opinion, 1, in every interaction with neutral agents.

Shifting Amplification Strategy. For comparison, we also assess a slightly more complex strategy in which manipulative agents initially send an opinion different from the target opinion, but shift the signal by some specified interval so that they arrive at the target opinion—1—within the 5,000 total time steps. We implement this strategy in three different ways: (1) manipulative agents start at opinion −1 and shift by 0.01 every

200 time steps; (2) manipulative agents start at opinion −0.4 and shift by 0.01 every 300 time steps; and (3) manipulative agents start at −0.1 and shift by 0.01 every 400 time steps.

3 Results

3.1 Assimilative Influence

The simple amplification strategy is sufficient for consistently moving the entire neutral population to the target opinion when assimilative influence is used. Without the presence of a strategy, the population forms a consensus near the average opinion from the agents' initial states. When the simple amplification strategy is used, this consensus still forms, but is then pulled to the target opinion (Fig. 1).

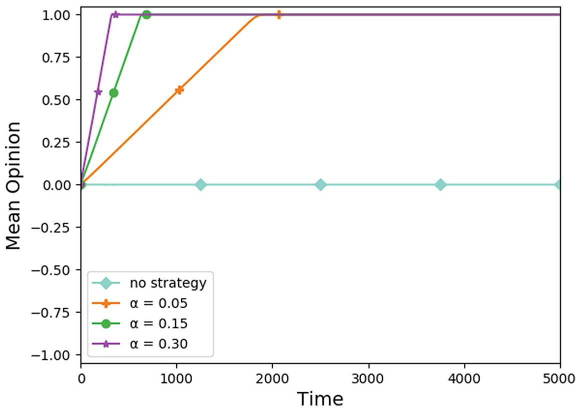


Fig. 1. Comparison of the mean opinion of the neutral population over time under assimilative influence using the simple amplification strategy with three different amplification values, α. Results displayed for N = 100 agents with initial opinions uniformly distributed across the opinion range. Under these conditions, increasing α results in the mean opinion of the population reaching the target opinion, 1, more rapidly.

3.2 Similarity-Biased Influence

Under the assumptions of similarity-biased influence, the simple amplification strategy is successful only for $\varepsilon = 0.9$ and with sufficient amplification. However, the three shifting strategies were able to move to the entire population to the target opinion in every simulation run (Fig. 2).

When the threshold, ε, is decreased to 0.6, the simple strategy never moves the entire population to the target population, though it can move some proportion of the population. Two of the three shifting strategies succeed reliably in moving the whole population to the target opinion. Notably, the third shifting strategy performs slightly worse on average as α is increased (Fig. 3).

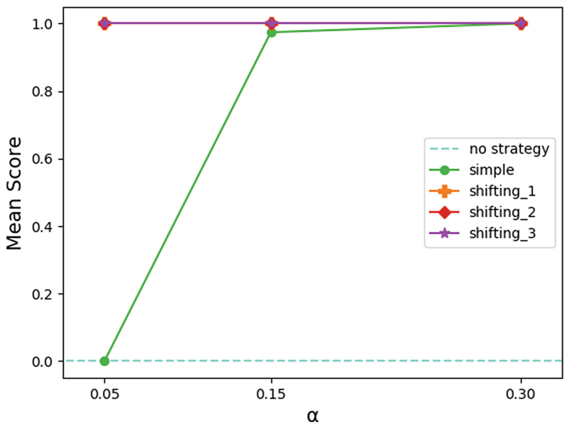


Fig. 2. Mean score comparison of the simple and shifting amplification strategies under the similarity-biased influence mechanism with ε = 0.9. Under these conditions, the simple amplification strategy is capable of moving the entire neutral population to the target opinion, but requires sufficient amplification to do so.

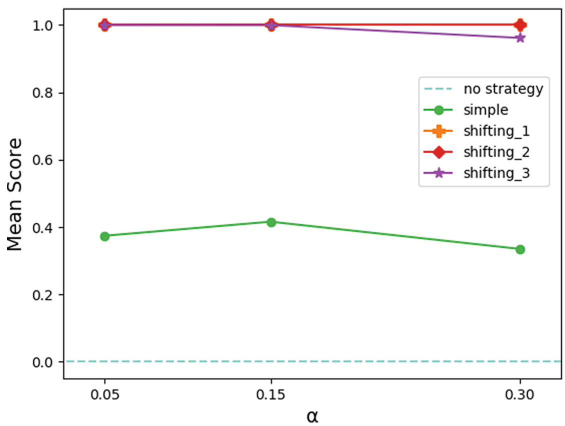


Fig. 3. Mean score comparison of the simple and shifting amplification strategies under the similarity-biased influence mechanism with ε = 0.6. Under this more restrictive threshold value, the average proportion of agents moved to the target opinion by the simple amplification strategy does not exceed 0.5. Increasing α does not always result in a higher score as seen when α is increased from 0.15 to 0.30 in the mean scores of the simple amplification strategy and the third shifting strategy.

3.3 Repulsive Influence

Under the repulsive influence mechanism, the neutral agent population bipolarizes and thus some proportion of the population reaches the target opinion even without the presence of an influence strategy. Here, the simple amplification strategy performs worse on average than using no strategy at all in moving the neutral population to the

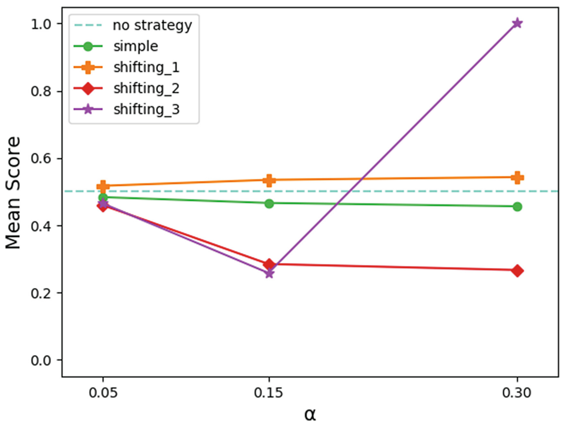


Fig. 4. Mean score comparison for each influence strategy when $\varepsilon = 0.9$ under the repulsive influence mechanism. Of the strategies analyzed, only the third shifting strategy is capable of reliably moving the entire neutral population to the target opinion, but does so only for $\alpha = 0.3$. This strategy also displays a nonlinear effect of increasing α: the mean score lowers when α is increased from 0.05 to 0.15, but then increases when α is further increased to 0.30.

target opinion. Increasing α lowers the mean score of the simple strategy for the two threshold values we explore (Figs. 4 and 5).

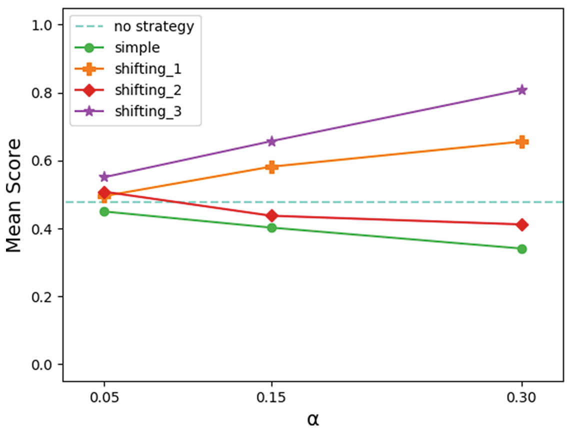


Fig. 5. Mean score comparison of the simple and shifting amplification strategies under the repulsive influence mechanism with $\varepsilon = 0.6$. None of the explored strategies succeed in moving the entire neutral population to the target opinion.

When $\varepsilon = 0.9$, only one of the shifting strategies we explore is capable of reliably moving the entire neutral population to the target opinion, though sufficient amplification is required for this to occur (Fig. 4). When ε is further decreased to 0.6, neither the simple nor shifting strategies are able to move the entire neutral population to the target opinion, though the first and third shifting strategies are capable of moving a majority of the population to the target opinion (Fig. 5).

4 Discussion

Our findings suggest that it is likely incorrect to assume that automated information campaigns are capable of causing a majority of individuals to adopt an opinion simply by artificially amplifying that opinion, except under fairly strict assumptions. It is crucially important to note that all of the results described above do not have a concrete relationship with actual automated information campaigns intended to manipulate actual human populations. However, these theoretical results do indicate a few useful points. The effectiveness of automated information campaigns, carried out by bots, is often assumed to follow simply from the repetition of opinions by bots. Yet, even in the highly simplified models we discuss, this repetition—corresponding to our simple amplification strategy—is only effective under extremely narrow conditions. As soon as any sort of criteria is introduced for influence to take place, whether in similarity-biased influence or in repulsive influence, the simple repetition of an opinion becomes much less reliable at affecting opinions or even completely ineffectual. Additionally, the variety of outcomes we see for many circumstances spells trouble for making accurate predictions about human populations. In other words, if it is difficult to predict outcomes for extremely simple agents (such as is the case for repulsive influence when $\varepsilon = 0.6$), it must certainly be difficult to do so for humans with all of the added complexity which would ensue.

This work suggests several next steps for future research. One such step is to experiment with the addition of counter-strategies, where one strategy might have 1 as its target opinion, while the other might have -1. Another step is to conduct similar experiments, but with the addition of network topologies governing which agents are able to interact with each other and to then assess influence strategies within various topologies. Additionally, redefining the goals of an influence strategy may be useful. For example, instead of a strategy having a target opinion, it may have a target distribution of opinions, such as bipolarization or a uniform diversity of opinions.

In conclusion, we have suggested that agent-based models can make a useful contribution to the study of automated information campaigns. They do so by forcing us to acknowledge the logical outcomes of our assumptions, and thus see how well-founded those assumptions are. Additionally, the ways in which these highly simplified models violate our common sense are likely to lead to ways in which models of opinion dynamics can be improved. Perhaps it seems too easy to manipulate the agents under assimilative influence because there is some key ingredient absent from the model, such as the perceived credibility of agents. Digging into these violations of our intuition will likely lead to more robust models of opinion dynamics generally.

Acknowledgements. This research is funded in part by the U.S. National Science Foundation (IIS-1636933, ACI-1429160, and IIS-1110868), U.S. Office of Naval Research (N00014-10-1-0091, N00014-14-1-0489, N00014-15-P-1187, N00014-16-1-2016, N00014-16-1-2412, N00014-17-1-2605, N00014-17-1-2675), U.S. Air Force Research Lab, U.S. Army Research Office (W911NF-16-1-0189), U.S. Defense Advanced Research Projects Agency (W31P4Q-17-C-0059) and the Jerry L. Maulden/Entergy Endowment at the University of Arkansas at Little Rock. The authors thank the three anonymous referees whose comments were immensely helpful in improving this paper.

References

1. Metaxas, P.T., Mustafaraj, E.: Social media and the elections. Science **338**(6106), 472–473 (2012)
2. Woolley, S.: Automating power: social bot interference in global politics. First Monday **21** (4) (2016)
3. Ferarra, E., Varol, O., Davis, C., Menczer, F., Flammini, A.: The rise of social bots. Commun. ACM **59**(7), 96–104 (2016)
4. Lazer, D.M.J., et al.: The science of fake news. Science **359**(6380), 1094–1096 (2018)
5. Flache, A., et al.: Models of social influence: towards the next frontiers. JASSS **20**(4), 2 (2017)
6. Deffuant, G., Neau, D., Amblard, F., Weisbuch, G.: Mixing beliefs among interacting agents. Adv. Complex Syst. **3**(1–4), 87–98 (2000)
7. Hegselmann, R., Krause, U.: Opinion dynamics and bounded confidence: models, analysis and simulation. JASSS **5**(3) (2002)
8. Hegselmann, R., König, S., Kurz, S., Niemann, C., Rambau, J.: Optimal opinion control: the campaign problem. JASSS **18**(3), 18 (2015)
9. Smaldino, P.E.: Models are stupid, and we need more of them. In: Vallacher, R.R., Nowak, A., Read, S.J. (eds.) Computational Social Psychology. Psychology Press (2017)
10. Edmonds, B.: Different modelling purposes. In: Edmonds, B., Meyer, R. (eds.) Simulating Social Complexity. Understanding Complex Systems. Springer, Cham (2017)

Implicit Learning and Creativity in Human Networks: A Computational Model

Marwa Shekfeh and Ali A. Minai(✉)

Department of Electrical Engineering and Computer Science,
University of Cincinnati, Cincinnati, OH 45221, USA
marwa.shekfeh@gmail.com, Ali.Minai@uc.edu

Abstract. With rare exceptions, new ideas necessarily emerge in the minds of individuals through the recombination of existing ideas, but the epistemic repertoire for this recombination is supplied largely by ideas the individual has acquired from external sources, including interaction with peers. When agents hear new ideas and integrate them into their minds, they also *implicitly* create potential new ideas which can then become explicit as new ideas through later introspection. In this research, we use a multi-agent model to study such implicit learning in a social network and its relationship with the number of unique novel ideas actually expressed by agents in the network. We focus on the impact of two crucial factors: (1) The structure of the social network; and (2) The selectivity of agents in accepting ideas from their peers. We look at both *latent ideas*, i.e., those that are still implicit in the minds of individual agents, and novel expressed ideas, i.e., those that are expressed for the first time in the network. The results show that both network structure and the selectivity of influence have significant impact on the outcomes – especially in a system with misinformation.

1 Introduction

It is widely recognized that new ideas often arise through recombination of parts of existing ideas. Since this process almost always occurs in the minds of individuals, it is driven by the content and organization of knowledge in these minds, which, in turn, is acquired from various sources, including peer such interaction also produces implicit or incidental learning that occurs as a side-effect [7,11]. This implicit knowledge comprising *latent ideas* is precisely the substrate from which novel ideas emerge through introspection.

Because this type of learning is very difficult to study experimentally, it is important to understand it through computational modeling. We have developed a computational model called the *Multi-Agent Network for the Implicit Learning of Associations* (MANILA) to study how implicit learning and the generation of novel conceptual combinations in a social network are influenced by network structure and the selectivity of agents in absorbing ideas from their peers.

A. J. Morales et al. (Eds.): ICCS 2018, SPCOM, pp. 147–154, 2018.
https://doi.org/10.1007/978-3-319-96661-8_15

MANILA is inspired not only by the attributes of social systems, but also of complex systems in general – especially biological evolution. In the model, ideas are described as combinations of elementary *concepts*. New ideas arise via recombination (and possibly mutation) of motifs from known ideas, much as new organisms arise through the recombination and mutation of genes from existing organisms. The minds of the agents in this system provide a medium for such recombination and mutation, with the agents' communications disseminating the resulting products as ideas. As in evolution, ideas are reinforced indirectly – in this case, by conferring higher merit status on their originators and communicators. Of course, the analogy is not exact, but it is in accord with standard models of innovation and creativity [2,3,5,8,10,14], and builds on computational models of ideation as an associative dynamical process developed recently by our research group [6,9]. The research we describe is closely related to the models of norm emergence in social networks [4,12,15].

2 System Description

The Model proposed in this paper is based on *cognitive and generative agents* engaged in the implicit acquisition of associative information through social interaction. The agents are connected in a *social network* whose structure can be controlled within the model. Each agent, i, has an *epistemic network* (EN), E^i, whose nodes represent *concepts* and edges indicate *associations* between concepts. Thus, The system is a network of networks. A quasi-clique (sufficiently connected sub-network) of the EN with between 6 and 10 nodes is considered an *idea*.

The Oracle: To provide a concrete definition of learning, we assume that there is a large and fixed *ideal epistemic network* (IEN) of concepts and associations known only to an *Oracle*, which serves as the representation of reality or the ground truth. The agents' initial ENs represent partial, noisy samples from the IEN. Whenever an agent generates an idea, it gets a *reward* from the Oracle based on the consistency of the generated idea with the IEN, but without any detailed feedback on which associations in the idea were correct or false. However, the *cumulative* (and time-discounted) reward of an agent represents its *fitness* value, and is visible to all agents.

The Social Network: The social network between agents is a symmetric adjacency matrix of connectivities $c_{ij} \in \{0, 1\}$ from A_j to A_i according to a specified *connectivity structure*. Four types of connectivity structures are considered: (1) *Random* (Erdös-Renyi) (RA); (2) *Modular* (MO), where agents are divided into K equal-sized *communities* with dense intra-community and sparse inter-community connectivity; (3) *Small-world* (SW), based on the small-world model of Watts and Strogatz [16]; and (4) *Scale-free* (SF), with a power law node degree distributed generated via the preferential attachment model by Barabasi and Albert [1].

The networks are generated so that the total number of nodes are identical in the four cases and the total number of edges almost identical. All edges (i, j) with $c_{ij} = 1$ are assigned with *weight* values φ_{ij}representing social connection, drawn independently from a fixed distribution. Since weights are constrained to be between 0 and 1, it is natural to use a Beta distribution where $a, b \geq 0$

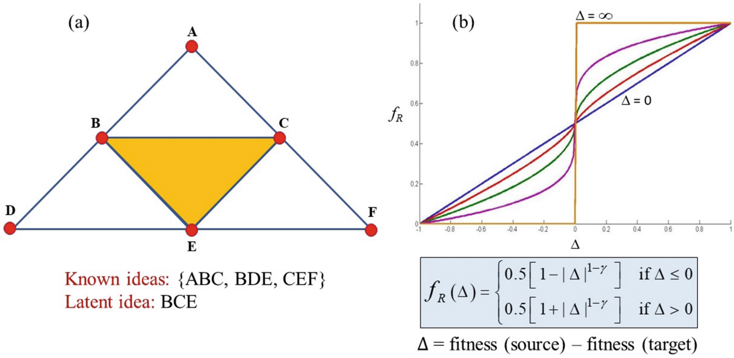
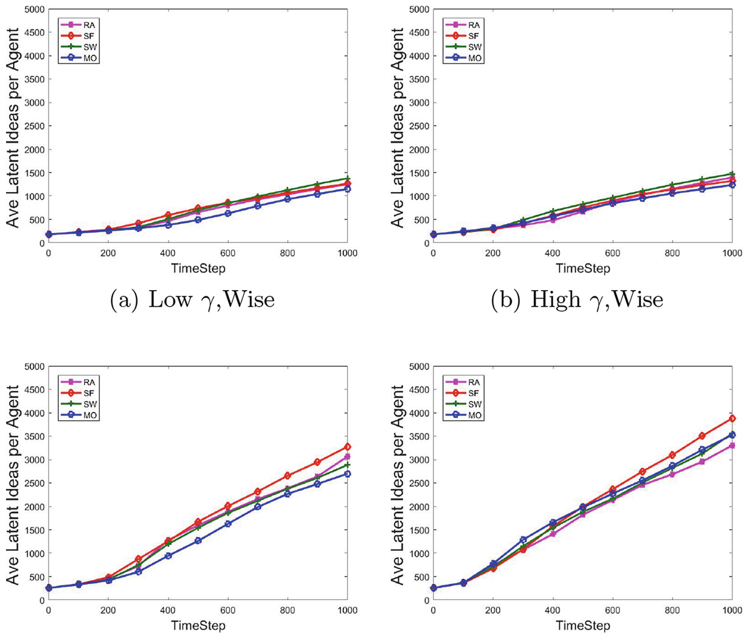


Fig. 1. (a) Latent idea: three known ideas lead to the formation of latent idea BCE. (b) The fitness selectivity function determines the probability of accepting an idea from a peer based on differential fitness, Δ

Agent Model: The knowledge of an agent A_i is defined by its Epistemic Network, $E^i(t)$, which is initialized as a partial, noisy sub-network of the Oracle's IEN. The noise has two aspects: Nodes and edges present in the IEN but not in E_i represent the *naivete* of A_i, i.e., concepts and associations that A_i does not know; edges absent in the IEN but present in E_i represent *misinformation*, or false associations. The knowledge base of any real individual always includes both true and false associations, and *normal agents* in MANILA are initialized with both, thus endowing each agent with some initial misinformation. However, as a limiting case, it is also interesting to consider *wise agents* that are initialized only with naivete but no misinformation (See [13] for details).

After initialization, the system goes through N iterations where each agent A_i performs the following actions: (1) It generates ideas from its EN and expresses a subset of its generated ideas based on its *expressiveness*; (2) It accumulates reward from the Oracle in response to its expressed ideas, thereby acquiring its fitness; (3) It receives ideas from agents in its peer group and assimilates them into its EN based on its general *receptivity*, r_i, and its specific *attentiveness*, Λ_{ij}, for the source agent, A_j. Thus, E_i changes with time through learning.

The implicit goal of learning, i.e., the way we evaluate system's performance, is for each agent to bring its EN closer to the IEN, i.e., learn more of the "truth". The agents are not aware of this explicitly, and see themselves as merely communicating with peers and assimilating some of their ideas. Very importantly, the agent ENs are initialized such that every concept and edge in the IEN is present



(a) Low γ,Wise (b) High γ,Wise

(c) Low γ,Normal (d) High γ,Normal

Fig. 2. Number of latent true ideas per agent with low selectivity (left column) and high selectivity (right column) in an All Wise population (top row) and Normal population (bottom row).

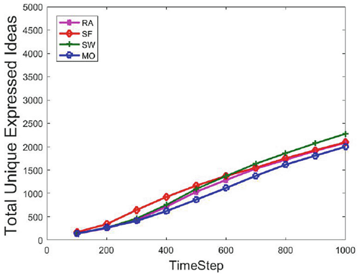
in the *union* of all initial ENs in the population, so that complete learning is possible at least in principle. Thus, the process of communication and learning inexorably reduces naivete, but it also spreads misinformation.

Idea Generation and Expression: Every agent periodically searches its EN for an idea to express via an attention-biased search process, and if an appropriate idea is found, expresses it to its peer group. At a given time, t, this process involves four steps:

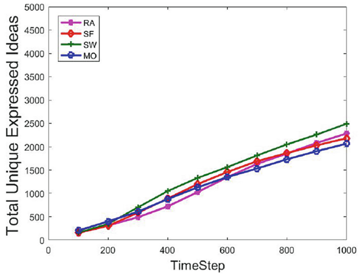
Step 1 - Specification of Attentional Focus: This step identifies the concepts and associations that agent A_i currently has in mind. This begins with a small subset, $S^i(t)$, of $E_i(t)$ nodes called the *seed set*. Next, all the nodes connected to $S^i(t)$ in the agent's EN are identified. The resulting sub-network, $E_i^B(t)$, of active nodes is called *attended epistemic network* (AEN) of the agent at time t.

Step 2 - Introspection: Generation of Potential Ideas: This step generates the set $G^i(t)$ of all possible *potential ideas*, $g_k^i(t)$, within the AEN, defined as connected subsets of between 6 and 10 nodes in its AEN.

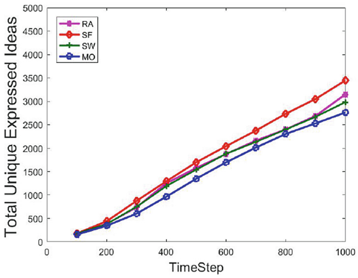
Step 3 - Evaluation of Potential Ideas: Once agent A_i has generated the set of potential ideas, $G^i(t) = \{g_k^i(t)\}$, it evaluates the *coherence*, $q_k^i(t)$, of each idea based on the density of its internal connectivity in its AEN. More densely



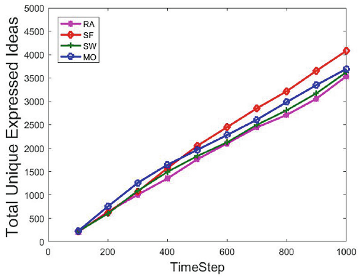
(a) Low γ,Wise



(b) High γ,Wise



(c) Low γ,Normal



(d) High γ,Normal

Fig. 3. Number of unique generated ideas with low selectivity (left column) and high selectivity (right column) in a All Wise population (top row) and Normal population (bottom row).

connected ideas are considered superior. Coherence values below a threshold ϕ are set to 0. The final coherence values are used via a roulette wheel mechanism to choose one idea for expression, so more coherent ideas have a higher probability of being expressed. A_i then broadcasts the chosen idea to all its immediate neighbors. If no potential idea meets the coherence threshold ϕ, the agent remains silent in that time-step.

Idea Evaluation and Reward Accumulation: Whenever an agent A_i expresses an idea $g^i(t)$, the idea is evaluated by the Oracle as follows:

$$\chi(g) = \frac{e_{true} - e_{false}}{e_{total}} \tag{1}$$

where e_{total} is the number of edges in g, e_{true} is the number of edges in g that are also present in the IEN, and e_{false} is the number of edges present in the g but not in the IEN. The reward is then calculated as:

$$R(\chi(g^i(t))) = \begin{cases} 0.5[1 - (\chi)^{1-\gamma}] & if\ \chi\ \leq\ 0 \\ 0.5[1 + (\chi)^{1-\gamma}] & if\ \chi\ >\ 0 \end{cases} \tag{2}$$

where λ is the *reward selectivity* parameter and $\chi(g)$ represents the Oracle's quality evaluation function for ideas. The received reward then causes the agent's fitness to be updated as:

$$\xi_i(t+1) = (1-\sigma)\xi_i(t) + \sigma R(\chi) \tag{3}$$

where $\sigma \in (0,1)$ is the *fitness adaptation rate* parameter and is set to a small value. Thus, the fitness, ξ_i of agent A_i is a non-specific cumulative indicator of the quality of its recently expressed ideas.

Idea Reception and Assimilation: The probability of agent A_i accepting an idea from agent A_j is: $P_{accept}(i,j) = r_i f_R(\Delta_{ij})$, where $r_i \in [0,1]$ is the general receptivity of agent A_i to ideas from others. and $f_R(\Delta_{ij})$ is the *fitness selectivity function*, which determines how attentiveness depends on the fitness differential $\Delta_{ij} = \xi_j(t) - \xi_i(t)$. It is shown in Fig. 1(b). If the *fitness selectivity* parameter is $\gamma = 0$, f_R is a linear function of differential fitness. As γ increases towards 1, $f_R(\Delta)$ approaches a threshold function at $\Delta = 0$, so the agent always accepts ideas from agents with higher fitness and never from those with lower fitness than itself.

Idea Assimilation: Once A_i has accepted an idea, g, from A_j, it must be assimilated into A_i's epistemic network. Any new nodes in the idea are simply added to E_i. Edges present (absent) in both E_i and g remain connected (disconnected); edges present in g but not in E_i are added; and those present in E_i but not in g are deleted with a very small probability (0.01 in simulations). Thus, the result is to stochastically make E^i more consistent with the received idea g. Both addition and removal of edges is possible, which is crucial for reducing misinformation. The current version of MANILA does not provide for the addition of misinformation.

Known and Latent Ideas: Every quasi-clique of 6 to 10 nodes in an agent's EN is a potential idea, but only a subset of these are considered *known ideas*, i.e., those that were used in creating its initial mindset, those it has generated through introspection, and those it has assimilated from others. The rest are considered *latent ideas*. As an agent thinks and learns, more ideas become known to it, but many more latent ideas are created to replace each one that becomes known. Figure 1(a) shows how latent ideas emerge inevitably from known ideas through the creation of novel quasi-cliques. Indeed, Latent ideas are the reservoir from which agents discover and generate novel ideas.

3 Results and Discussion

The simulations reported in this paper focus on two classes of ideas: (1) The average number of *correct* latent ideas per agent at time t (results in Fig. 2); and (2) The cumulative total number of *all* unique novel ideas expressed in the system up to time t (results in Fig. 3). Note that only latent ideas consistent with the Oracle are considered in (1), but unique expressed ideas may be of varying quality. The unique ideas represent the *creativity* of the population. Aside from the four types of social networks, two types of populations are considered: (1) *All Wise*, where there none of the agents are initially misinformed; and (2) *Normal*, where agents initially have a moderate amount of misinformation. In each case, we consider two cases of agent selectivity: (1) *High* γ (= 0.9), where agents learn almost exclusively from those with higher fitness; and (2) *Low* γ (= 0.1), where some learning from lower fitness agents is also possible. Each data-point is averaged over ten independent runs.

To interpret the results of Fig. 2 (implicit learning), consider that agent fitness values are initialized to the fraction of misinformation in the agents' initial ENs, so agents in the All Wise population all begin with a fitness of 1. Also, since their ideas exclusively comprise correct edges their fitness remains at 1, which means that $\Delta_{ij} = 0$ and thus $f_R(\Delta_{ij}) = 0.5$ for all A_i and A_j throughout regardless of γ. Thus, it is reasonable that γ should make no difference to the amount of implicit learning in the All Wise case, as is seen in Fig. 2(a) and (b). Agents in the Normal population, in contrast, begin with a broader distribution of fitnesses, which widens as agents express ideas of varying quality. Higher γ, i.e., higher preference for learning from fitter peers, then leads to filtering out the learning of false associations (on average), making the number of latent ideas learned higher, as is seen in Fig. 2(c) and (d). Figure 2(c), (d) also indicate that the SF (scale-free) social architecture is somewhat superior for spreading latent ideas in a Normal population. This highlights the role of hub agents in these populations, who can quickly disseminate true (and false) associations to large numbers of peers and serve as shortcuts between less well-connected peers.

Figure 3 (unique expressed ideas) shows much more diverse results. The high and low γ cases are still quite similar for the All Wise population (plots (a) and (b)). Network architecture makes only a small difference, but the SW (small-world) architecture has increasingly superior performance with time. The results for the Normal population show a remarkable amount of variation. The most significant feature of the graphs is the emergence of a "breakout" phenomenon where creativity picks up suddenly, leading to a higher total of ideas.

At both levels of selectivity, the SF structure clearly leads to the greatest amplification of creativity, which reflects the fact that this structure also produces the most implicit learning. However, it should be noted that the rapid dissemination of misinformation on a SF social network presumably fuels both good and bad novel ideas, and the latter may be a significant part of the creative boom. The RA and MO networks take longer to trigger breakouts, which is reasonable, since information spreads more slowly on these than on the SF network. The γ value also makes a significant difference, with higher γ leading to more creativity on average for all but the SW case. However, it is not clear whether the early breakout of MO in the low γ case and that of RA in the high γ case is significant. For some reason, the SW network does not show a breakout of creativity, though one could emerge later. There is also no difference in the performance of the SW case for low and high γ. It is interesting to note that the SW network did not lag the RA or MO networks in implicit learning.

Another interesting observation is that, for some network structures, the All Wise population is more creative than the Normal population, even though the former has less latent learning and the ideas expressed by it are constrained to all be correct.

Acknowledgment. This work was supported in part by National Science Foundation INSPIRE grant BCS-1247971 to Ali Minai.

References

1. Barabasi, A.L., Albert, R.: Emergence of scaling in random networks. Science **286**, 509–511 (1999)
2. Brown, V., Tumeo, M., Larey, T., Paulus, P.: Modeling cognitive interactions during group brainstorming. Small Group Res. **29**, 495–526 (1998)
3. Campbell, D.T.: Blind variation and selective retention in creative thought as in other knowledge processes. Psychol. Rev. **67**, 380–400 (1960)
4. Delgado, J., Pujol, J.M., Sangüesa, R.: Emergence of coordination in scale-free networks. Web Intell. Agent Syst. **1**, 131–138 (2003)
5. Fauconnier, G., Turner, M.: The Way We Think: Conceptual Blending and the Mind's Hidden Complexities. Basic Books, New York (2003)
6. Iyer, L.R., Doboli, S., Minai, A.A., Brown, V.R., Levine, D.S., Paulus, P.B.: Neural dynamics of idea generation and the effects of priming. Neural Netw. **22**, 674–686 (2009)
7. Lewicki, P., Czyzewska, M., Hoffman, H.: Unconscious acquisition of complex procedural knowledge. J. Exp. Psychol. Learn. Mem. Cogn. **13**(4), 523–530 (1987)
8. Mednick, S.: The associative basis of the creative process. Psychol. Rev. **69**(3), 220–232 (1962)
9. Minai, A.A., Iyer, L.R., Padur, D., Doboli, S.: A dynamic connectionist model of idea generation. In: Proceedings of the IJCNN 2009, pp. 2109–2116 (2009)
10. Nijstad, B.A., Stroebe, W.: How the group affects the mind: a cognitive model of idea generation in groups. Pers. Soc. Psychol. Rev. **3**, 186–213 (2006)
11. Reber, A.S.: Implicit learning of artificial grammars. J. Verbal Learn. Verbal Behav. **6**(6), 855–863 (1967)
12. Sen, O.: Effects of social network topology and options on norm emergence. In: Coordination, Organizations, Institutions and Norms in Agent Systems. LNCS, vol. 6069, pp. 211–222 (2010)
13. Shekfeh, M.: MANILA: a multi-agent framework for emergent associative learning and creativity in social networks. Ph.D. thesis, Department of Electrical Engineering and Computer Science, University of Cincinnati (2017)
14. Simonton, D.K.: Scientific creativity as constrained stochastic behavior: the integration of product, person, and process perspectives. Psychol. Bull. **129**, 475–494 (2003)
15. Villatoro, D., Sen, S., Sabater-Mir, J.: Topology and memory effect on convention emergence. In: IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technologies, pp. 233–240 (2009)
16. Watts, D., Strogatz, S.: Collective dynamics of "small-world" networks. Nature **393**, 440–442 (1998)

On the Shadow Moments of Apparently Infinite-Mean Phenomena

Nassim Nicholas Taleb[1(✉)] and Pasquale Cirillo[2]

[1] Tandon School of Engineering, New York University, New York, USA
NNT1@nyu.edu

[2] Applied Probability Group, Delft University of Technology, Delft, Netherlands

Abstract. We propose an approach to compute the conditional moments of fat-tailed phenomena that, only looking at data, could be mistakenly considered as having infinite mean. This type of problems manifests itself when a random variable Y has a heavy-tailed distribution with an extremely wide yet bounded support.

We introduce the concept of dual distribution, by means of a logarithmic transformation that smoothly removes the upper bound. The tail of the dual distribution can then be studied using extreme value theory, without making excessive parametric assumptions, and the estimates one obtains can be used to study the original distribution and compute its moments by reverting the transformation.

The central difference between our approach and a simple truncation is in the smoothness of the transformation between the original and the dual distribution, allowing use of extreme value theory.

Keywords: Power laws · Complex networks · Econophysics

1 Introduction

Consider a heavy-tailed random variable Y with finite support $[L, H]$. W.l.o.g. set $L >> 0$ for the lower bound, while for upper one H, assume that its value is remarkably large, yet finite. It is so large that the probability of observing values in its vicinity is extremely small, so that in data we tend to find observations only below a certain $M << H < \infty$.

Figure 1 gives a graphical representation of the problem. For our random variable Y with remote upper bound H the real tail is represented by the continuous line. However, if we only observe values up to $M << H$, and - willing or not - we ignore the existence of H, which is unlikely to be seen, we could be inclined to believe the tail is the dotted one, the apparent one. The two tails are indeed essentially indistinguishable for most cases, as the divergence is only evident when we approach H.

A. J. Morales et al. (Eds.): ICCS 2018, SPCOM, pp. 155–164, 2018.
https://doi.org/10.1007/978-3-319-96661-8_16

Now assume we want to study the tail of Y and, since it is fat-tailed and despite $H < \infty$, we take it to belong to the so-called Fréchet class[1]. In extreme value theory [3], a distribution F of a random variable Y is said to be in the Fréchet class if $\bar{F}(y) = 1 - F(y) = y^{-\alpha}L(y)$, where $L(y)$ is a slowly varying function. In other terms, the Fréchet class is the class of all distributions whose right tail behaves as a power law.

Looking at the data, we could be led to believe that the right tail is the dotted line in Fig. 1, and our estimation of α shows it be smaller than 1. Given the properties of power laws, this means that $E[Y]$ is not finite (as all the other higher moments). This also implies that the sample mean is essentially useless for making inference, in addition to any considerations about robustness [6]. But if H is finite, this cannot be true: all the moments of a random variable with bounded support are finite.[2]

A solution to this situation could be to fit a parametric model, which allows for fat tails and bounded support, such as for example a truncated Pareto [5]. But what happens if Y only shows a Paretian behavior in the upper tail, and not for the whole distribution? Should we fit a mixture model?

In the next section we propose a simple general solution, which does not rely on strong parametric assumptions.

2 The Dual Distribution

Instead of altering the tails of the distribution we find it more convenient to transform the data and rely on distributions with well known properties. In Fig. 1, the real and the apparent tails are indistinguishable to a great extent. We can use this fact to our advantage, by transforming Y to remove its upper bound H, so that the new random variable Z - the dual random variable - has the same tail as the apparent tail. We can then estimate the shape parameter α of the tail of Z and come back to Y to compute its moments or, to be more exact, to compute its excess moments, the conditional moments above a given threshold, view that we will just extract the information from the tail of Z.

Take Y with support $[L, H]$, and define the function

$$\varphi(Y) = L - H \log\left(\frac{H - Y}{H - L}\right). \tag{1}$$

We can verify that φ is "smooth": $\varphi \in C^{\infty}$, $\varphi^{-1}(\infty) = H$, and $\varphi^{-1}(L) = \varphi(L) = L$. Then $Z = \varphi(Y)$ defines a new random variable with lower bound L and an infinite upper bound. Notice that the transformation induced by $\varphi(\cdot)$ does not depend on any of the parameters of the distribution of Y.

---

[1] Note that treating Y as belonging to the Fréchet class is a mistake. If a random variable has a finite upper bound, it cannot belong to the Fréchet class, but rather to the Weibull class [2].

[2] Note that we require that the upper bound H be both finite and known –as with the world's population or the capital of a bank. In case H is not known, we fail to see the difference with an unbounded random variable.

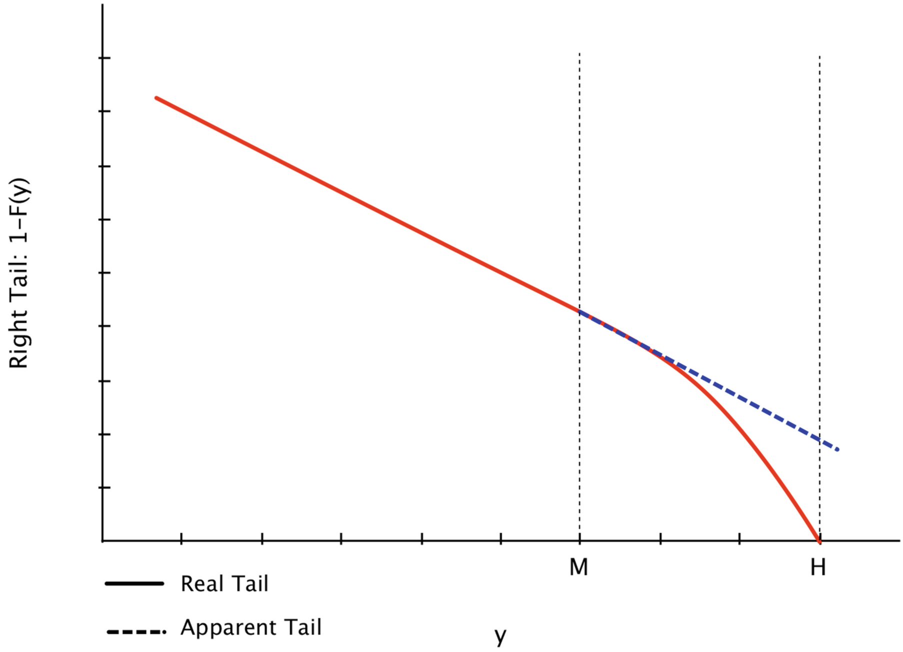


Fig. 1. Graphical representation of what may happen if one ignores the existence of the finite upper bound H, since only M is observed.

By construction, $z = \varphi(y) \approx y$ for very large values of H. This means that for a very large upper bound, unlikely to be touched, the results we get for the tail of Y and $Z = \varphi(Y)$ are essentially the same, until we do not reach H. But while Y is bounded, Z is not. Therefore we can safely model the unbounded dual distribution of Z as belonging to the Fréchet class, study its tail, and then come back to Y and its moments, which under the dual distribution of Z could not exist.[3]

The tail of Z can be studied in different ways, see for instance [3,4]. Our suggestions is to rely on the so-called de Pickands, Balkema and de Haan's Theorem [2]. This theorem allows us to focus on the right tail of a distribution, without caring too much about what happens below a given threshold u. In our case $u \geq L$.

Consider a random variable Z with distribution function G, and call G_u the conditional df of Z above a given threshold u. We can then define the r.v. W, representing the rescaled excesses of Z over the threshold u, so that

$$G_u(w) = P(Z - u \leq w | Z > u) = \frac{G(u+w) - G(u)}{1 - G(u)},$$

[3] Note that the use of logarithmic transformation is quite natural in the context of utility.

for $0 \leq w \leq z_G - u$, where z_G is the right endpoint of G.

Pickands, Balkema and de Haan have showed that for a large class of distribution functions G, and a large u, G_u can be approximated by a Generalized Pareto distribution, i.e. $G_u(w) \rightarrow GPD(w;\xi,\sigma),$ as $u \rightarrow \infty$ where

$$GPD(w;\xi,\sigma) = \begin{cases} 1-(1+\xi\frac{w}{\sigma})^{-1/\xi} & if\, \xi \neq 0 \\ 1-e^{-\frac{w}{\sigma}} & if\, \xi = 0 \end{cases}, \; w \geq 0. \tag{2}$$

The parameter ξ, known as the shape parameter, and corresponding to $1/\alpha$, governs the fatness of the tails, and thus the existence of moments. The moment of order p of a Generalized Pareto distributed random variable only exists if and only if $\xi < 1/p$, or $\alpha > p$ [3]. Both ξ and σ can be estimated using MLE or the method of moments [2].[4]

3 Back to Y: The Shadow Mean

With f and g, we indicate the densities of Y and Z.

We know that $Z = \varphi(Y)$, so that $Y = \varphi^{-1}(Z) = (L-H)e^{\frac{L-Z}{H}} + H$.

Now, let's assume we found $u = L^* \geq L$, such that $G_u(w) \approx \text{GPD}(w;\xi,\sigma)$. This implies that the tail of Y, above the same value L^* that we find for Z, can be obtained from the tail of Z, i.e. G_u.

First we have

$$\int_{L^*}^{\infty} g(z)\,\mathrm{d}z = \int_{L^*}^{\varphi^{-1}(\infty)} f(y)\,\mathrm{d}y. \tag{3}$$

And we know that

$$g(z;\xi,\sigma) = \frac{1}{\sigma}\left(1+\frac{\xi z}{\sigma}\right)^{-\frac{1}{\xi}-1}, \qquad z \in [L^*,\infty). \tag{4}$$

Setting $\alpha = \xi^{-1}$, we get

$$f(y;\alpha,\sigma) = \frac{H\left(1+\frac{H(\log(H-L)-\log(H-y))}{\alpha\sigma}\right)^{-\alpha-1}}{\sigma(H-y)}, \;\; y \in [L^*,H], \tag{5}$$

or, in terms of distribution function,

$$F(y;\alpha,\sigma) = 1-\left(1+\frac{H(\log(H-L)-\log(H-y))}{\alpha\sigma}\right)^{-\alpha}. \tag{6}$$

[4] There are alternative methods to face finite (or concave) upper bounds, i.e., the use of tempered power laws (with exponential dampening) [7] or stretched exponentials [8]; while being of the same nature as our exercise, these methods do not allow for immediate applications of extreme value theory or similar methods for parametrization.

Clearly, given that φ is a one-to-one transformation, the parameters of f and g obtained by maximum likelihood methods will be the same—the likelihood functions of f and g differ by a scaling constant.

We can derive the shadow mean[5] of Y, conditionally on $Y > L^*$, as

$$E[Y|Y > L^*] = \int_{L^*}^{H} y\, f(y; \alpha, \sigma)\, \mathrm{d}y, \tag{7}$$

obtaining

$$E[Y|Z > L^*] = (H - L^*)e^{\frac{\alpha\sigma}{H}} \left(\frac{\alpha\sigma}{H}\right)^{\alpha} \Gamma\left(1-\alpha, \frac{\alpha\sigma}{H}\right) + L^*. \tag{8}$$

The conditional mean of Y above $L^* \geq L$ can then be estimated by simply plugging in the estimates $\hat{\alpha}$ and $\hat{\sigma}$, as resulting from the GPD approximation of the tail of Z. It is worth noticing that if $L^* = L$, then $E[Y|Y > L^*] = E[Y]$, i.e. the conditional mean of Y above Y is exactly the mean of Y.

Naturally, in a similar way, we can obtain the other moments, even if we may need numerical methods to compute them.

Our method can be used in general, but it is particularly useful when, from data, the tail of Y appears so fat that no single moment is finite, as it is often the case when dealing with operational risk losses, the degree distribution of large complex networks, or other econophysical phenomena.

For example, assume that for Z we have $\xi > 1$. Then both $E[Z|Z > L^*]$ and $E[Z]$ are not finite[6]. Figure 1 tells us that we might be inclined to assume that also $E[Y]$ is infinite - and this is what the data are likely to tell us if we estimate $\hat{\xi}$ from the tail[7] of Y. But this cannot be true because $H < \infty$, and even for $\xi > 1$ we can compute the expected value $E[Y|Z > L^*]$ using Eq. (8).

Value-at-Risk and Expected Shortfall

Thanks to Eq. (6), we can compute by inversion the quantile function of Y when $Y \geq L^*$, that is

$$Q(p; \alpha, \sigma, H, L) = e^{-\gamma(p)} \left(L^* e^{\frac{\alpha\sigma}{H}} + He^{\gamma(p)} - He^{\frac{\alpha\sigma}{H}}\right), \tag{9}$$

where $\gamma(p) = \frac{\alpha\sigma(1-p)^{-1/\alpha}}{H}$ and $p \in [0, 1]$. Again, this quantile function is conditional on Y being larger than L^*.

From Eq. (9), we can easily compute the Value-at-Risk (VaR) of $Y|Y \geq L^*$ for whatever confidence level. For example, the 95% VaR of Y, if Y represents operational losses over a 1-year time horizon, is simply $VaR^Y_{0.95} = Q(0.95; \alpha, \sigma, H, L)$.

[5] We call it "shadow", as it is not immediately visible from the data.

[6] Remember that for a GPD random variable Z, $E[Z^p] < \infty$ iff $\xi < 1/p$.

[7] Because of the similarities between $1 - F(y)$ and $1 - G(z)$, at least up until M, the GPD approximation will give two statistically undistinguishable estimates of ξ for both tails [3].

Another quantity we might be interested in when dealing with the tail risk of Y is the so-called expected shortfall (ES), that is $E[Y|Y > u \geq L^*]$. This is nothing more than a generalization of Eq. (8).

We can obtain the expected shortfall by first computing the mean excess function of $Y|Y \geq L^*$, defined as

$$e_u(Y) = E[Y - u|Y > u] = \frac{\int_u^\infty (u-y) f(y;\alpha,\sigma)\mathrm{d}y}{1 - F(u)},$$

for $y \geq u \geq L^*$. Using Eq. (5), we get

$$e_u(Y) = (H-L)e^{\frac{\alpha\sigma}{H}} \left(\frac{\alpha\sigma}{H}\right)^\alpha \left(\frac{H \log\left(\frac{H-L}{H-u}\right)}{\alpha\sigma} + 1\right)^\alpha \times$$
$$\Gamma\left(1-\alpha, \frac{\alpha\sigma}{H} + \log\left(\frac{H-L}{H-u}\right)\right). \tag{10}$$

The Expected Shortfall is then simply computed as

$$E[Y|Y > u \geq L^*] = e_u(Y) + u.$$

As in finance and risk management, ES and VaR can be combined. For example we could be interested in computing the 95% ES of Y when $Y \geq L^*$. This is simply given by $VaR^Y_{0.95} + e_{VaR^Y_{0.95}}(Y)$.

4 Comparison to Other Methods

There are three ways to go about explicitly cutting a Paretan distribution in the tails (not counting methods to stretch or "temper" the distribution).

(1) The first one consists in hard truncation, i.e. in setting a single endpoint for the distribution and normalizing. For instance the distribution would be normalized between L and H, distributing the excess mass across all points.
(2) The second one would assume that H is an absorbing barrier, that all the realizations of the random variable in excess of H would be compressed into a Dirac delta function at H –as practiced in derivative models. In that case the distribution would have the same density as a regular Pareto except at point H.
(3) The third is the one presented here.

The same problem has cropped up in quantitative finance over the use of truncated normal (to correct for Bachelier's use of a straight Gaussian) vs. logarithmic transformation (Sprenkle [9]), with the standard model opting for logarithmic transformation and the associated one-tailed lognormal distribution. Aside from the additivity of log-returns and other such benefits, the models do not produce a "cliff", that is an abrupt change in density below or above, with the instability associated with risk measurements on non-smooth function.

As to the use of extreme value theory, Breilant et al. [1] go on to truncate the distribution by having an excess in the tails with the transformation $Y^{-\alpha} \rightarrow (Y^{-\alpha} - H^{-\alpha})$ and apply EVT to the result. Given that the transformation includes the estimated parameter, a new MLE for the parameter α is required. We find issues with such a non-smooth transformation. The same problem occurs as with financial asset models, particularly the presence an abrupt "cliff" below which there is a density, and above which there is none. The effect is that the expectation obtained in such a way will be higher than ours, particularly at values of $\alpha < 1$, as seen in Fig. 2.

We can demonstrate the last point as follows. Assume we observe distribution is Pareto that is in fact truncated but treat it as a Pareto. The density is $f(x) = \frac{1}{\sigma}\left(\frac{x-L}{\alpha\sigma}+1\right)^{-\alpha-1}, x \in [L, \infty)$. The truncation gives $g(x) = \frac{\left(\frac{x-L}{\alpha\sigma}+1\right)^{-\alpha-1}}{\sigma(1-\alpha^{\alpha}\sigma^{\alpha}(\alpha\sigma+H-L)^{-\alpha})}, x \in [L, H]$.

Moments of order p of the truncated Pareto (i.e. what is seen from realizations of the process), $M(p)$ are:

$$M(p) = \alpha e^{-i\pi p}(\alpha\sigma)^{\alpha}(\alpha\sigma - L)^{p-\alpha} \frac{\left(B_{\frac{H}{L-\alpha\sigma}}(p+1, -\alpha) - B_{\frac{L}{L-\alpha\sigma}}(p+1, -\alpha)\right)}{\left(\frac{\alpha\sigma}{\alpha\sigma+H-L}\right)^{\alpha} - 1} \tag{11}$$

where $B(.,.)$ is the Euler Beta function, $B(a,b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)} = \int_0^1 t^{a-1}(1-t)^{b-1}\,dt$.

We end up with $r(H, \alpha)$, the ratio of the mean of the soft truncated to that of the truncated Pareto.

$$r(H,\alpha) = e^{-\frac{\alpha}{H}}\left(\frac{\alpha}{H}\right)^{\alpha}\left(\frac{\alpha}{\alpha+H}\right)^{-\alpha}\left(\frac{\alpha+H}{\alpha}\right)^{-\alpha} \frac{\left(-\left(\frac{\alpha+H}{\alpha}\right)^{\alpha} + H + 1\right)}{(\alpha-1)\left(\left(\frac{\alpha}{H}\right)^{\alpha} - \left(\frac{\alpha+H}{H}\right)^{\alpha}\right)E_{\alpha}\left(\frac{\alpha}{H}\right)} \tag{12}$$

where $E_{\alpha}\left(\frac{\alpha}{H}\right)$ is the exponential integral $e_{\alpha}z = \int_1^{\infty} \frac{e^{t(-\alpha)}}{t^n}\,dt$.

Comparison to Weibull. One could then ask what is the advantage of our approach with respect to the more standard Weibull limiting class in extreme value theory [3]. After all, in both situations, the upper bound H is finite. As for all statistical models, the difference and the advantage lie in the intended use.

The estimation of H in the Weibull class requires not only H to be finite (as we also do), but also that, heuristically, H is not too far from the maximum one may observe in the data. That's why, for example, the Weibull class is a perfect tool to study the limits of human life span [11], where the usual maxima (people about 120 years old) can be assumed to be close to the actual upper bound (about 130–140 years).

In our case, as also stated at the beginning, H is definitely finite but also extremely far away in the tails, well beyond the usual maxima one can observe.

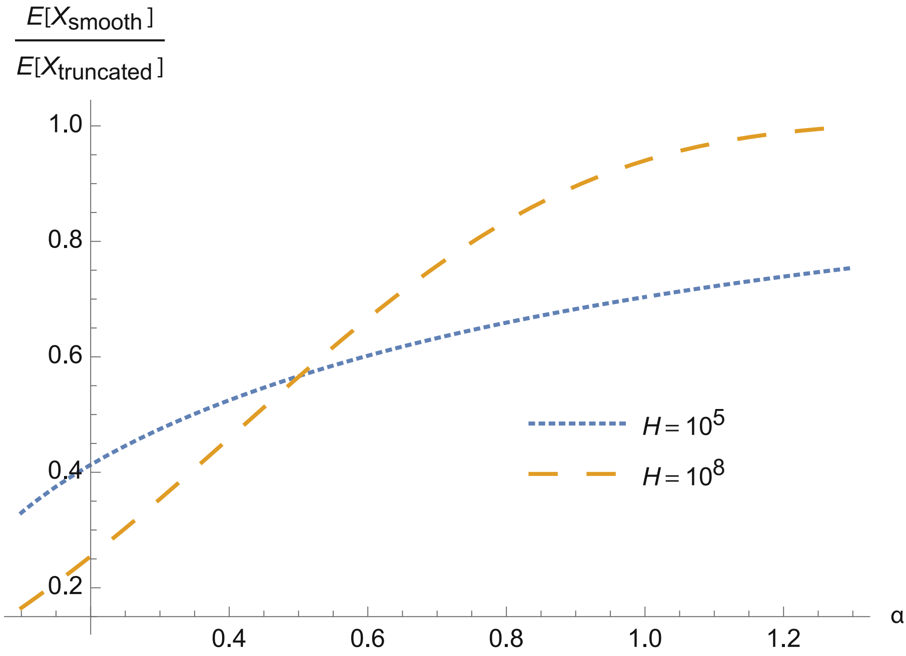


Fig. 2. Ratio of the expectation of smooth transformation to truncated.

WW2 killed millions of people, between 60 and 80 million depending on the estimates, but this number is for sure far away from the theoretical maximum an armed conflict could reach, also given the evolution of the military technology.

In extreme value statistics, in applications, when the upper bound is extremely large yet finite, one usually [3] prefers to use the Fréchet class to model the tail, accepting the risk of having an infinite mean, an event that is read as a sign of the unreliability and of the erratic nature of data. But this, in our opinion, is not only incorrect from a epistemological point of view, given that one knows the large upper bound and yet ignores this information, but also from a practical point of view, because ones shirks the responsibility of computing something that can be actually computed, even if done with all the caveats and uncertainties.

5 Applications

We run a list of proposed immediate and straightforward applications:

Operational Risk. The losses for a firm are bounded by the capitalization, with well-known maximum losses.

Capped Reinsurance Contracts. Reinsurance contracts almost always have caps (i.e., a maximum claim); but a reinsurer can have many such contracts on the same source of risk and the addition of the contract pushes the upper bound in such a way as to cause larger potential cumulative harm.

Violence. While wars are extremely fat-tailed, the maximum effect from any such event cannot exceed the world's population.

Credit Risk. A loan has a finite maximum loss, in a way similar to reinsurance contracts.

City Size. While cities have been shown to be Zipf distributed, the size of a given city cannot exceed that of the world's population.

Environmental Harm. While these variables are exceedingly fat-tailed, the risk is confined by the size of the planet (or the continent on which they take place) as a firm upper bound.

Complex Networks. The number of connections is finite.

Company Size. The sales of a company is bound by the GDP.

Earthquakes. The maximum harm from an earthquake is bound by the energy.

Hydrology. The maximum level of a flood can be determined.

References

1. Beirlant, J., Alves, I.F., Gomes, I., Meerschaert, M.: Extreme value statistics for truncated Pareto-type distributions (2014). arXiv preprint: arXiv:1410.4097
2. de Haan, L., Ferreira, A.: Extreme Value Theory: An Introduction. Springer, New York (2006)
3. Embrechts, P., Klüppelberg, C., Mikosch, T.: Modelling Extremal Events. Springer, Heidelberg (2003)
4. Falk, M., Hüsler, J., Reiss, R.-D.: Laws of Small Numbers: Extremes and Rare Events. Birkhäuser, Basel (2004)
5. Inmaculada, B.A., Meerschaert, M.M., Panorska, A.K.: Parameter estimation for the truncated Pareto distribution. J. Am. Stat. Assoc. **101**, 270–277 (2006)
6. Maronna, R., Martin, R.D., Yohai, V.: Robust Statistics - Theory and Methods. Wiley, New York (2006)
7. Rachev, S.T., Kim, Y.S., Bianchi, M.L., Fabozzi, F.J.: Financial Models with Lévy Processes and Volatility Clustering. Wiley, Hoboken (2011)

8. Laherrere, J., Sornette, D.: Stretched exponential distributions in nature and economy: "fat tails" with characteristic scales. Eur. Phys. J. B Condens. Matter Complex Syst. **2**(4), 525–539 (1998)
9. Sprenkle, C.: Warrant prices as indicators of expectations and preferences. Yale Econ. Essays **1**, 179–231 (1961)
10. Taleb, N.N.: Incerto Technical Companion. The Statistical Consequences of Fat Tails, vol. 1 (2018). www.fooledbyrandomness.com
11. Einmahl, J., Einmahl, J., de Haan, L.: Limits to Human Life Span Through Extreme Value Theory. Center Discussion Paper Series No. 2017-051. CentER, Center for Economic Research, Tilburg (2017)

A Lexical Network Approach for Identifying Suicidal Ideation in Clinical Interview Transcripts

Ulya Bayram[1(✉)], Ali A. Minai[1], and John Pestian[2]

[1] Department of Electrical Engineering and Computer Science, University of Cincinnati, Cincinnati, OH 45221, USA
bayramua@mail.uc.edu, Ali.Minai@uc.edu

[2] Department of Pediatrics and Cincinnati Children's Hospital Medical Center, University of Cincinnati, 3333 Burnet Avenue, Cincinnati, OH 45229, USA
john.pestian@cchmc.org

Abstract. Preventing suicide requires early identification of suicidal ideation. In this research, we propose an approach to evaluate whether an individual's statements during a clinical interview can be classified as coming from a suicidal or non-suicidal mindset. To do so, we compare the statements with distinct lexical associative networks constructed from corpora of suicidal and control texts. Each node in these networks is a word, and the weight of the edge between every word pair indicates how strongly the words are associated in that corpus. Several metrics of association are evaluated in this work. Preliminary results show good classification performance with above 75% accuracy on novel test data.

Keywords: Lexical networks · Suicidal ideation · Suicide
Clinical interviews · Word association · Spreading activation
Classification

1 Introduction

Worldwide, there are approximately 800,000 suicides a year, and at least six people are devastated by each of these deaths [15]. In the United States, suicide is among the top ten leading causes of death [10]. Most methods for detecting suicidal risk rely on the identification of suicidal ideation via self-reported answers to direct questions [5,23], which can be inaccurate [26] because individuals often try to hide suicidal thoughts [29].

This has motivated the search for other clues, such as objective suicidal thought, or behavioral markers [29]. Various data collections obtained from individuals with suicidal ideation have allowed researchers to use common supervised machine learning methods and experiment with some standard natural language processing (NLP) techniques. Some of these studies have shown state-of-the-art classification rates within their datasets, but their real-world application performance is yet to be determined. Many of these approaches have a common

A. J. Morales et al. (Eds.): ICCS 2018, SPCOM, pp. 165–172, 2018.
https://doi.org/10.1007/978-3-319-96661-8_17

drawback: they do not account for the cognitive aspect of the problem which is a key element in predicting suicide [28]. In this study, we describe an approach that accounts for this cognitive element.

During self-expression, thoughts and ideas are organized into sentences [14], and the use of specific words in the same sentence is a strong indicator that the corresponding concepts are associated in the individual's mind. We use this to build *lexical networks* from interview transcripts to capture the pattern of association between concepts in the suicidal and non-suicidal mindsets and determine which mindset a given expression is more consistent with. Experimental results show that the method detects suicidal ideation with greater than 75% accuracy for novel expressions, i.e. those not used to build the reference lexical networks. As a followup, we also experiment with spreading activation [1,33] in the lexical network to further enrich the discriminative process. This allows expressions to be evaluated not only based on their content but also on related thoughts. Results show some initial promise for improving the potential for identifying suicidal ideation, though this remains to be validated. The approach we describe is not restricted by the suicidal ideation detection problem, and can be applied to any other binary text-classification problem where reference corpora are available.

2 Method

2.1 Data

This study uses two prospectively assembled data collections containing interviews conducted with volunteering subjects, all native English speakers, by a mental health professional asking five simple questions regarding the subject having any fears, feeling hurt, having hopes, feeling angry, and having secrets. The first collection includes 60 transcripts of interviews conducted with adolescents [30], where half are patients admitted to the Cincinnati Children's Hospital Medical Center (CCHMC) for suicidal ideation or attempts and the rest are the patients admitted for orthopedic injuries with no recent or past suicidal ideation. The second data collection extends the first. It includes interviews collected from patients admitted to one of three hospitals: CCHMC, University of Cincinnati Medical Center (UC), and Princeton Community Hospital (PCH) who were either suicidal, mentally ill, or a member of the control group [32]. In contrast with the first collection, this study includes follow-up interviews conducted a month later, mentally ill patients, and adults. Since we are interested in detecting suicidal ideation, we exclude the mentally ill patient interviews; we merge the two collections and get a final dataset with 470 transcribed and annotated interviews. To standardize data for analysis, we use standard NLP techniques: converting each character to lowercase, removing punctuation while keeping only the end-of-sentence information, removing the English language stop words available in the Natural Language Toolkit (NLTK) [7], and also removing infrequent words with fewer than 5 occurrences.

2.2 Lexical Networks

Building networks from words and word associations is one approach for semantic memory research [1,9], and in research problems aiming to discover the semantic relatedness between words in the lexical domain [8,11,12,16,17,19,27,34]. In this study, we use this approach to identify the thought-patterns models of suicidal and non-suicidal individuals. This method is based on the idea that thoughts emerge from a substrate of associations between concepts [6,18,20,21], and that the analysis of associations between words in the statements of an individual can provide access to the semantics of their thinking [4,12,16,17,25].

We initialize both lexical networks (suicidal and control) using the respective training corpora by adding the pre-processed words spoken by suicidal ($V_{suicidal}$) and non-suicidal ($V_{control}$) interviewees as nodes. Since thoughts are organized and expressed through sentences [14], we use sentence-level word associations to establish connections between the words in the networks. However, NLP literature contains various word association metrics [11,12,22]. In our approach, we experiment with three of such metrics. The joint probability metric $p(i, j)$ represents the probability of observing words i and j together in the same sentence in the training corpus. The correlation coefficient metric, $cc(i, j)$, is a normalized covariance measure in the occurrence of i and j, and uses the in-sentence occurrence probabilities (p_i, p_j) of the words (see Eq. 1). Finally, the point-wise mutual information metric (PMI), $pmi(i, j)$, measures the statistical dependence in the occurrence of i and j (see Eq. 2). We use a normalized version of the PMI metric.

$$cc(i, j) = \frac{p(i, j) - p_i p_j}{\sqrt{p_i(1 - p_i)p_j(1 - p_j)}} \tag{1}$$

$$pmi(i, j) = log\big(\frac{p_i p_j}{p(i, j)}\big)/log(p(i, j)) \tag{2}$$

Following the computation of word associations from training corpora, we set the negative values of the correlation coefficient and PMI meaning non-association to zero weights. Thus, non-associated word pairs remain unconnected in the networks and the positive edge weights indicate the connection strengths between words. Samples from the two networks are shown in Fig. 1, where the differences between the suicidal and control networks are visually evident.

2.3 Evaluating the Fitness of Thoughts

To determine whether a given novel statement is more consistent with a suicidal or non-suicidal mindset, we take the words from the statement and extract the sub-networks of each reference lexical network with only these words. An example of the sub-network obtained by matching the words in a short test statement with those in the suicidal lexical network with joint probabilities as edge weights is shown in Fig. 2a. We hypothesize that the strength of connectivity within the two sub-networks obtained in this way can identify which lexical network the test statement fits better in, thus determining whether the underlying statement

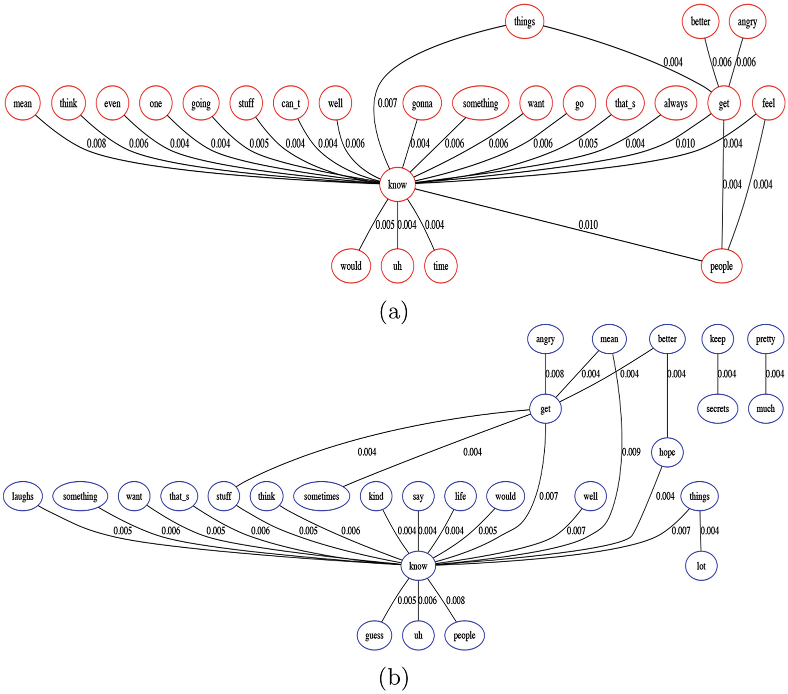


Fig. 1. The strongest connections in joint probability networks of suicidal and control training corpora obtained by thresholding the edge weights, returned (a) out of 48, 169 edges in the suicidal network, (b) out of 38, 132 edges in the control network.

is more consistent with a suicidal mindset or a non-suicidal one. To do this, we define a *fitness function* that is applied to the extracted sub-network, E:

$$\Phi(E) = d_e^S(E) - d_e^C(E) \tag{3}$$

$$d_e^G(E) = d^E - d^G = \frac{\sum_{i,j \in E} w_{i,j}}{n_E(n_E - 1)/2} - \frac{\sum_{i,j \in G} w_{i,j}}{n(n-1)/2} \tag{4}$$

where S and C superscripts are the suicidal and control networks, G is the reference lexical network (suicidal or control), n is the number of nodes in G, n_E is the number of nodes in sub-network E, and $d_e^G(E)$ is the *excess weight density* of E with respect to G, which quantifies the degree to which the strength of connectivity within E exceeds the value that would be expected for a random sub-network of the same size extracted from G. Such measures are commonly used in network studies [13, 24].

2.4 Spreading Activation

Spreading activation (SA) theory was developed by Quillan [33] as an attempt to model the associative nature of human semantic memory. Although there are many spreading activation methods [1–3, 31], the main idea is based on assigning initial activation values to nodes, and by some rules, letting the activation spread to the rest of the network via associative edges. To limit spreading, all spreading activation approaches include a loss factor that can be a decay function or a simple threshold parameter. Spreading activation is a directed process whereby active nodes activate those to which they are connected with sufficiently strong edges.

When a test expression is presented to the system, nodes corresponding to its words are activated, and this activation is allowed to spread until the limiting process stops the activation. All nodes activated in this episode then become part of the sub-network that is tested for fitness.

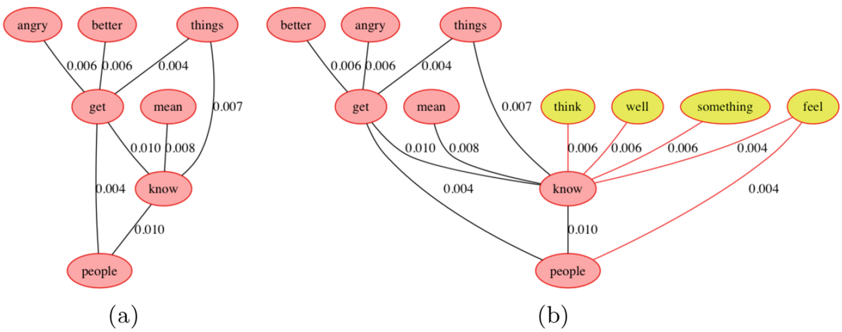


Fig. 2. Sample sub-networks obtained (a) without using spreading activation, (b) using spreading activation, where the nodes and connections added by spreading are highlighted.

As a result of the SA process, we obtain two sub-networks comprising the initial sub-networks plus words and connections activated by the spreading process. Thus, these sub-networks can be considered enriched or augmented versions of the original ones. Sample sub-networks obtained without SA and with it are shown in Fig. 2a and b respectively, with nodes added by the spreading activation results highlighted. The fitness of the augmented sub-networks is evaluated as with the original ones.

3 Results

For experiments, we increased the effective number of training and test cases by randomly and independently choosing ten test-training sets in $1/3-2/3$ ratio from both the suicidal and control corpora, training and testing separately on each set, and aggregating the results, giving a total of 750 test data points for each corpus.

The first experiment evaluated the performance of the method without spreading activation. A test data item was labeled as suicidal or control based on which lexical network gave it a higher fitness. The results obtained from 1,500 test cases are summarized in Table 1, where the best association metric is the correlation coefficient, with above 75% accuracy. Having such a high accuracy in this difficult task indicates that lexical networks are a promising method for doing such evaluations.

Table 1. Suicidal ideation detection results measured by classification accuracy and area under the ROC curve (AUC) obtained using three association metrics on 1,500 test cases.

Association metric	Accuracy	AUC score ± stdev
Joint probability	64.3%	74.8 % ± 2.5%
Point-wise Mutual Information (PMI)	73.6%	79.3% ± 2.3%
Correlation coefficient	**75.6%**	**81.6%** ± 2.1%

The second experiment uses spreading activation (SA) to evaluate the effect of the augmented sub-networks on detecting suicidal ideation. The addition of more words and connections to the sub-networks by spreading activation results an accuracy of 74.5% using the lexical networks built by the correlation coefficient metric, which is actually a very small *decrease* in accuracy over the original method. However, the potential benefits of the approach become clearer by looking at the number of true positives (TP) and false positives (FP), where positive indicates the suicidal case. Comparing TP and FP values between results obtained with (TP = 555, FP = 187) and without spreading activation (TP = 550, FP = 166) shows that having spreading activation allows an increase in TP rates, but at the cost of an increase in control cases being identified as suicidal. In this life-critical application, however, this may well be an acceptable trade-off.

4 Conclusion

In this study, we have proposed a lexical network approach to detecting suicidal ideation in self-expression data. Experiments on a novel dataset with actual interviews have shown that the approach is able to classify suicidal and non-suicidal ideation with high accuracy, which demonstrates the success of the approach not only for this application, but also for other binary classification problems involving textual data. The use of spreading activation as a part of the approach is potentially useful because associations play a key role in the organization of thought. However, the benefit remains to be demonstrated in the specific case of suicidal ideation analysis. Experiments have provided evidence that using spreading activation can improve suicidal detection rates at the cost of including more false positives. Since identifying suicidal individuals is far more important than potentially misidentifying some non-suicidal individuals as suicidal, the benefits of spreading activation may outweigh its cost. Also, while the data used was based on a careful identification of suicidal and non-suicidal individuals, it could not possibly control for every expression. Even non-suicidal individuals can occasionally express ideas consistent with a suicidal mindset, and vice versa. Thus, classifying expressions as a way of identifying an individual's entire mindset is inherently subject to some degree of uncertainty, which can only be alleviated by using more extensive data from each individual.

The methods described in this study represent only an initial phase of our research. The fact that even the simple methods described above give a high degree of discrimination indicates that they can serve as the basis for more sophisticated methods, which we will report on in the future.

Acknowledgment. We are grateful for the support of the University of Cincinnati College of Medicine, and the Department of Electrical Engineering and Computer Science. We are also grateful for the support of the Cincinnati Children's Hospital Medical Center, and its division of Biomedical Informatics.

References

1. Anderson, J.R.: A spreading activation theory of memory. J. Verbal Learn. Verbal Behav. **22**(3), 261–295 (1983)
2. Anderson, J.R.: Act: a simple theory of complex cognition. Am. Psychol. **51**(4), 355 (1996)
3. Anderson, J.R., Bothell, D., Byrne, M.D., Douglass, S., Lebiere, C., Qin, Y.: An integrated theory of the mind. Psychol. Rev. **111**(4), 1036 (2004)
4. Bales, M.E., Johnson, S.B.: Graph theoretic modeling of large-scale semantic networks. J. Biomed. Inform. **39**(4), 451–464 (2006)
5. Beck, A.T., Kovacs, M., Weissman, A.: Assessment of suicidal intention: the scale for suicide ideation. J. Consult. Clin. Psychol. **47**(2), 343 (1979)
6. Benedek, M., Jauk, E., Fink, A., Koschutnig, K., Reishofer, G., Ebner, F., Neubauer, A.C.: To create or to recall? Neural mechanisms underlying the generation of creative new ideas. NeuroImage **88**, 125–133 (2014)
7. Bird, S., Klein, E., Loper, E.: Natural Language Processing with Python. O'Reilly Media, Sebastopol (2009)
8. Budanitsky, A., Hirst, G.: Evaluating Wordnet-based measures of lexical semantic relatedness. Comput. Linguist. **32**(1), 13–47 (2006)
9. Collins, A.M., Loftus, E.F.: A spreading-activation theory of semantic processing. Psychol. Rev. **82**(6), 407 (1975)
10. Coppersmith, G., Leary, R., Whyne, E., Wood, T.: Quantifying suicidal ideation via language usage on social media. In: Joint Statistics Meetings Proceedings, Statistical Computing Section. JSM (2015)
11. De Deyne, S., Storms, G.: Word associations: network and semantic properties. Behav. Res. Methods **40**(1), 213–231 (2008)
12. Doumit, S., Marupaka, N., Minai, A.A.: Thinking in prose and poetry: a semantic neural model. In: Proceedings of IJCNN 2013 (2013)
13. Frank, K.A.: Identifying cohesive subgroups. Soc. Netw. **17**(1), 27–56 (1995)
14. Frege, G.: The thought: a logical inquiry. Mind **65**(259), 289–311 (1956)

15. Geneva: World Health Organization: Preventing suicide: A resource for media professionals. Technical report, World Health Organization (2017)
16. Kenett, Y., Kenett, D., Ben-Jacob, E., Faust, M.: Global and local features of semantic networks: evidence from the Hebrew mental lexicon. PLoS ONE **6**, e23912 (2011)
17. Kenett, Y., Anaki, D., Faust, M.: Investigating the structure of semantic networks in low and high creative persons. Front. Hum. Intell. **8**, 407 (2014)
18. Landauer, T., Dumais, S.: A solution to plato's problems: the latent semantic analysis theory of acquisition, induction and representation of knowledge. Psychol. Rev. **104**, 211–240 (1997)
19. Marupaka, N., Iyer, L.R., Minai, A.A.: Connectivity and thought: the influence of semantic network structure in a neurodynamical model of thinking. Neural Netw. **32**, 147–158 (2012)
20. McRae, K., de Sa, V., Seidenberg, M.: On the nature and scope of featural representations of word meaning. J. Exp. Psychol. Gen. **126**, 99–130 (1997)
21. Mednick, S.: The associative basis of the creative process. Psychol. Rev. **69**(3), 220–232 (1962)
22. Mei, M., Vanarase, A., Minai, A.A.: Chunks of thought: finding salient semantic structures in texts. In: Proceedings of IJCNN 2014 (2014)
23. Miller, I.W., Norman, W.H., Bishop, S.B., Dow, M.G.: The modified scale for suicidal ideation: reliability and validity. J. Consult. Clin. Psychol. **54**(5), 724 (1986)
24. Moody, J., White, D.R.: Structural cohesion and embeddedness: a hierarchical concept of social groups. Am. Soc. Rev. **68**, 103–127 (2003)
25. Motter, A.E., de Moura, A.P.S., Lai, Y.C., Dasgupta, P.: Topology of the conceptual network of language. Phys. Rev. E **65**, 065102 (2002)
26. Murray, D.: Is it time to abandon suicide risk assessment? Br. J. Psychiatry Open **2**(1), e1–e2 (2016). https://doi.org/10.1192/bjpo.bp.115.002071
27. Nelson, D.L., Xu, J.: Effects of implicit memory on explicit recall: set size and word frequency effects. Psychol. Res. **57**, 203–214 (1995)
28. Neuringer, C.: The thinking processes in suicidal women. In: Why Women Kill Themselves, pp. 43–52 (1988)
29. Nock, M.K., Park, J.M., Finn, C.T., Deliberto, T.L., Dour, H.J., Banaji, M.R.: Measuring the suicidal mind: implicit cognition predicts suicidal behavior. Psychol. Sci. **21**(4), 511–517 (2010)
30. Pestian, J.P., Grupp-Phelan, J., Bretonnel Cohen, K., Meyers, G., Richey, L.A., Matykiewicz, P., Sorter, M.T.: A controlled trial using natural language processing to examine the language of suicidal adolescents in the emergency department. Suicide Life Threat. Behav. **46**(2), 154–159 (2015)
31. Pestian, J.P., Matykiewicz, P., Duch, W., Glauser, T.A., Kowatch, R.A., Grupp-Phelan, J.M., Sorter, M.: Processing text with domain-specific spreading activation methods. US Patent 9,477,655 (2016)
32. Pestian, J.P., Sorter, M., Connolly, B., Bretonnel Cohen, K., McCullumsmith, C., Gee, J.T., Morency, L.P., Scherer, S., Rohlfs, L.: A machine learning approach to identifying the thought markers of suicidal subjects: a prospective multicenter trial. Suicide Life Threat. Behav. **47**(1), 112–121 (2017)
33. Quillan, M.R.: Semantic memory. Technical report, DTIC Document (1966)
34. Steyvers, M., Tenenbaum, J.B.: The large-scale structure of semantic networks: statistical analyses and a model of semantic growth. Cogn. Sci. **29**(1), 41–78 (2005)

Applying Complexity Science with Machine Learning, Agent-Based Models, and Game Engines: Towards Embodied Complex Systems Engineering

Michael D. Norman(✉), Matthew T. K. Koehler, Jason F. Kutarnia, Paul E. Silvey, Andreas Tolk, and Brittany A. Tracy

The MITRE Corporation, Bedford, MA 01730, USA
mnorman@mitre.org

Abstract. The application of Complexity Science, an undertaking referred to here as Complex Systems Engineering, often presents challenges in the form of agent-based policy development for bottom-up complex adaptive system design and simulation. Determining the policies that agents must follow in order to participate in an emergent property or function that is not pathological in nature is often an intensive, manual process. Here we will examine a novel path to agent policy development in which we do not manually craft the policies, but allow them to emerge through the application of machine learning within a game engine environment. The utilization of a game engine as an agent-based modeling platform provides a novel mechanism to develop and study intelligent agent-based systems that can be experienced and interacted with from multiple perspectives by a learning agent. In this paper we present results from an example use-case and discuss next steps for research in this area.

Keywords: Artificial intelligence · Complexity · Emergence Reinforcement learning

1 Introduction

The application of the science of complexity, specifically attempts at engineering complex adaptive systems (CAS) [9], present the scientist or systems engineer with the non-trivial task of crafting agent policies that achieve the desired emergent properties of the system, assuming they are known. Complex Systems Engineering [8] addresses this new set of challenges, but requires a significantly extended set of methods, many of them from the domain of simulation, and in particular agent-based approaches which can be enriched by artificial intelligence methods [20]. We can use these machine learning (ML) techniques to effectively derive agent-based policies that maximize the probability of the desired

A. J. Morales et al. (Eds.): ICCS 2018, SPCOM, pp. 173–183, 2018.
https://doi.org/10.1007/978-3-319-96661-8_18

emergent effects (via feedback of information on agent decisions during training) while minimizing the probability of emergent outcomes we deem negative, which we refer to as pathological emergence.

The focus of this work is on the application of reinforcement learning (RL) to evolve and engineer collections of agents well-suited to producing and maintaining an emergent property of interest. Ongoing work cross-pollinating the fields of agent-based modeling and ML that these authors are aware of are focused on utilizing ML to derive agent policies from existing datasets [19], whereas in this work we utilize a game engine to enable an embodied RL paradigm [14] based solely on experience gained in the model of the complex system itself.

The recent integration of Unity [22], a popular game engine, with TensorFlow, an open source machine learning framework [18] has created an opportunity to build agent-based models that can learn from their own embodied experience. Our preliminary work here has enabled us to plot a course for using this technology to employ imitation learning from observing human-driven embodied experiences within their world, as well as new depths of human-machine teaming applied to agent-based modeling through true embodiment of the human inside of the model using virtual reality (VR). The implications for the integration of the CAS modeling field and the serious gaming field appear on the surface to be potentially far-reaching, including training autonomous vehicles [9], and perhaps virtual avatars capable of passing the Turing test.

In what follows, we explore the applicability of such a technology stack to designing an agent-based search and rescue model. Development of polices and results will be discussed. Future work will explore the impact of further agent heterogeneity, scaling, and higher-fidelity environments.

2 Technological Enablers

Real-time game engines such as Unity and Unreal [3], which are capable of producing life-like experiences, high-definition head-mounted displays (HMDs) and controllers which are designed to bring hand and body presence into virtual environments such as the HTC Vive [5], and consumer-grade graphical processing units (GPUs) designed for high-performance parallel processing such as those made by NVIDIA [10], have put the power to create agents trained by human interaction in virtual worlds in the average computer enthusiast's hands. Real-time generated virtual worlds could become ideal training grounds for the development of strategies for learning and problem solving in both virtual and real-world environments, as evidenced by the investment being made in companies who are using game-like environments to train artificial intelligence (AI) for real world autonomous car deployment [1].

Our work here represents an initial first step in this direction for the Applied Complexity Science community, and demonstrates the feasibility of agent-based modeling in a high-fidelity game environment combined with an RL mechanism. Future work on imitation learning may leverage alternative learning processes as appropriate.

3 Agent Policy Development

Agent-based modeling policies are challenging to approach for many reasons, not the least of which is the sensitivity of complex systems to their initial conditions [15], which means a successful policy must be robust enough to withstand latent, unpredictable, and path-dependent edge-cases. This leads a CAS designer or modeler to seek absolute minimalism in their policy implementations in order to avoid generating inexplicable pathological emergence; from simple rules complexity emerges. In this paper we present a different method of developing such robust policies via the application of ML techniques within a game engine environment. The results of a search-and-rescue use-case implemented using the Unity game engine and TensorFlow ML framework are discussed.

3.1 Determining Agent Policies When Modeling Complex Adaptive Systems

One of the best examples of a well-known and human-understandable agent policy set is that of the boids algorithm [11]. Through tuning of the simple rules of attraction, repulsion, and alignment, a flock of birds emerges. Previous work leveraging emergent properties of the boids rule-set has shown the emergence of a flock is an embodied phenomenon which depends greatly on context and scale, as well as very minimalist tweaking of the known, presumably 'stable' rule sets [9]. By stable, we mean capable of producing an emergent stability at the scale of the system. For example, a flock does not cease to be a flock just because there is instability among the population. This is sometimes known as metastability [6], but for our purposes, stability is an adequate term. It is also worth noting that a truly well-engineered CAS would demonstrate the properties of antifragility [16], but this is a lofty goal and will remain unobtainable until the science of complexity is able to catch its tools up to its theory. This paper is but a small attempt to move forward on that front.

If we are willing to accept a certain level of opacity in our agent policy sets and we are interested in problems of embodiment, then the creation of a capability to develop agent-based policies via ML in a 3D game engine is an excellent means of exploring the emergent dynamics of a complex system.

3.2 Reinforcement Learning

RL has a long history as a means of developing learning agents, and some work has even focused on the emergence of cooperation in relatively low-dimensional systems such as Pong [17], as well as within our own experimental framework [14], but its use in the development of agent-based modeling policies for studying and simulating CAS on a larger scale is still quite nascent, with only hints of the implications envisioned thus far [4].

The agents are trained via an RL implementation, so they proceed through many iterations of the simulation before training is considered complete. Training terminates after a certain number of time-steps. We proceed to repeat this process a number of times and compare outputs. We are then able to insert the newly-trained agent into the simulation to confirm it functions as expected. Selecting which trained agent to instantiate as multiple agents in the system (assuming homogeneous agents) in order to achieve a desired aggregate behavior is currently a manual process.

Five agents simultaneously inhabit the environment, interacting with one another and contributing to the learning of a single shared policy during each training run; the policy also controls each of the agents. The agent policy initially starts out doing very random things, and learns optimal rules through reinforcement-based trial-and-error.

3.3 Proximal Policy Optimization

The RL algorithm implemented at the core of the reinforcement learners is known as Proximal Policy Optimization (PPO). PPO was developed and released by OpenAI [12], and is their default algorithm for RL.

3.4 Performance

The measure of performance we will focus on for this use-case is simply Cumulative Reward at the conclusion of training. We will consider the cumulative reward across all agents at the conclusion of a given policy's training run to be the measure of emergent system performance, despite the fact that it is literally the sum of its parts. We can assume that a stable policy has been learned when the cumulative reward is positive. More interesting measures of emergence will be applied in future work.

4 Technology Stack

For this effort, Unity3D version 2017.3.1f1 was used in combination with TensorFlow version 1.6.0 and TensorFlowSharp version 1.6.0-pre1. Anaconda 5.1 and Python 3.6 were used to create a virtual environment and run Unity's external training API, respectively. The open source ml-agents package [23] from Unity Technology was the critical piece that integrated all of these technologies. Early pilot work involved training with discrete NVIDIA GTX 1080 GPUs, but these are optional. The ml-agents package was used to train a TensorFlow neural network graph from the RL experience of embodied agents situated in a game world.

5 Search and Rescue Use-Case

The open-source ml-agents package [23] was adapted to emulate autonomous search and rescue in a contested environment. In our simulation, the agents were cast as autonomous helicopters attempting to collect soldiers within a contested environment. A screen capture of the training environment is shown in Fig. 1. We implemented constrained peripheral vision for the agents as we consider it to be a rudimentary emulation of partial observability. With five agents acting and effecting concurrently in the simulation, the presence of multiple autonomous helicopter agents in the same "airspace" seeking the same basic objective was considered analogous to a dynamic multi-agent rescue situation in a theater of battle or other contested environment containing bad actors. Finally, the unpredictable distribution and periodic redistribution of soldiers within the rescue area required the agents to learn responsive behaviors tailored to stochastic conditions that cannot be anticipated in any actionable sense, only responded to prudently (positively rewarding) or imprudently (negatively rewarding): rampant exploration, but also care when maneuvering to a rescue so as not to encounter hostiles. To establish a baseline understanding of helicopter agent behavior, as well as to demonstrate the technological method first and foremost, we modeled five helicopters endeavoring to rescue friendly soldiers and avoid hostile soldiers, but a logical extension of this work could include dividing helicopters into discrete teams, positively rewarding agents when they rescue affiliated soldiers and negatively rewarding them for encounters with hostile soldiers. We populated the model with 20 friendly and 20 hostile soldiers. However, for every friendly soldier rescued, another is promptly respawned (this is not the case for hostile soldiers). The game area is reset at regular intervals to 20 and 20 soldiers of each category, and also randomly reassigns origin positions to the autonomous rescue helicopters.

We executed 10 experimental runs of our Search and Rescue application for 50,000 time steps.

In pilot runs, the initial reward structure (+1 for friendlies and −1 for hostiles) was not sufficient to produce the desired behavior of "colliding" with friendly soldiers and avoiding encounters with hostile soldiers. The helicopters' form entails that they often collide with hostile soldiers unintentionally from the side while they are still on a lateral trajectory (the hostiles are out-of-sight in the agents' periphery). In these initial trials, helicopters learned to remain stationary because they were frequently, though incidentally, negatively rewarded when they collided with clusters of bad soldiers.

In response, the reward structure was modified to negatively reward the helicopters with −1 points for every time step when they were not colliding with a soldier, in order to discourage inactivity. At this point, the helicopters learned to fly in lateral translation across the rescue area, thereby colliding with as many objects for +1 and −1 rewards as possible (either benefiting by +1 points or experiencing a −1 reward, the same result as if they had remained static). To discourage this lateral translation strategy, the reward structure was modified to negatively reward collision with a wall (on the bounds of the rescue area) at −100 points. Finally, the reward for collecting friendlies was increased to +500 points to offset the −1 reward increment applied at every non-collision timestep. Instead of proportional negative reward for collecting hostiles, the value was reduced to only one-fifth of the positive reward, −100 points. Finally, the simulation had

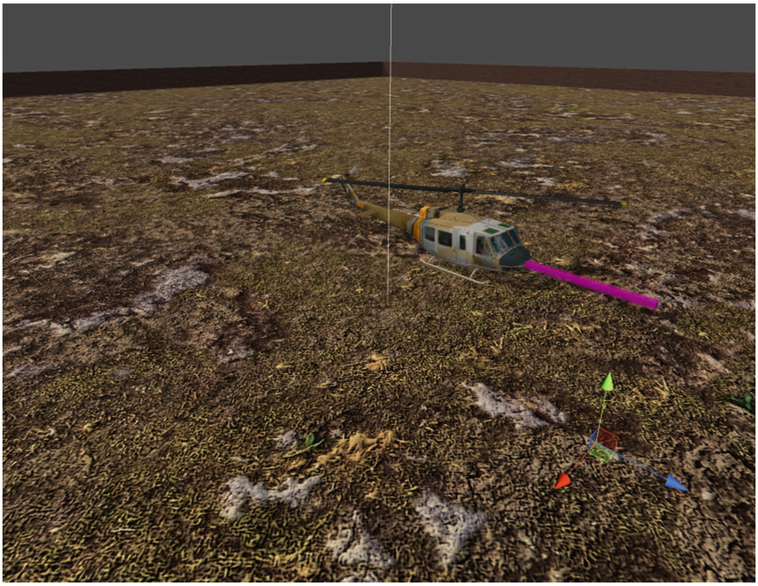


Fig. 1. Training environment screen capture.

achieved a form in which the reward structure was truly conducive to helicopters learning to explore the game area and attempting to rescue friendly soldiers.

The hyperparameter settings used for this experiment can be found in Table 1. These were informed by the recommended defaults of the ml-agents package.

6 Results and Discussion

Interesting summary results from the training can be found in Figs. 2 and 3, which show the Cumulative Reward and Entropy, respectively, across all runs.

Learners emerged that were capable of performing the task at hand, but whose entropy levels were still quite high at the termination of the training round that created them. We interpret this result as follows: these agents, although having constructed successful policies for survival, had also created a strategy which involved continual exploration. The implications of a novelty-seeking agent for a search and rescue use case seem rather tangible, as this 'injection of randomness' [16] may lead to nudging an emergent system dynamic out of an unsuccessful attractor's area of influence. For a summary of dynamical systems theory in context of agent-to-agent influence see Liebovitch et al. [7]. The varying levels of entropy across runs that was exhibited at the conclusion of training is indicative of the pervasiveness of exploration as a fruitful strategy. In fact the agent policy with highest total cumulative reward (as shown in the yellow line of Fig. 2) did not have the lowest entropy at the conclusion of its training, as can be seen in the dark blue line of Fig. 3.

Table 1. Hyperparameters.

Hyperparameter	Value
Batch size	1024
Beta	5.00E-03
Buffer size	10240
Epsilon	0.2
Gamma	0.99
Hidden units	128
Lambd	0.95
Learning rate	3.00E-04
Max steps	5.00E+04
Memory size	256
Normalize	FALSE
Num epoch	3
Num layers	2
Time horizon	64
Summary freq	1000
Use recurrent	FALSE

7 Future Work

The general need for research agendas to support better support of system of systems challenges [21] and complex systems engineering [2] has been identified and builds the general frame for this section, which provides particular research topics derived from the machine learning and agent-based metaphor applied in the work presented in this paper.

7.1 Hyperparameter Tuning

Hyperparameter tuning is a fundamental part of machine learning and must be carefully considered if one hopes to construct a high-performing autonomous agent. There are a number of intelligent algorithms that could be incorporated into our training workflow but investigation into the most effective was outside the scope of this paper. Recent research in simulation calibration may be applicable to support these ideas in future efforts [24].

The naïve approach is a simple grid search, but this breaks down when there are numerous parameters to tune or the model's sensitivity is nonlinear with respect to the hyperparameter, which is the case with neural networks. This can be overcome by using a logarithmically-scaled grid where sensitive areas of the hyperparameter space are heavily sampled. Other more advanced methods which attempt to minimize the number of trials necessary to converge on the optimum parameter values include Bayesian optimization and Sequential Model-based Algorithm Configuration. Future work will include more robust hyperparameter exploration.

7.2 Apply Machine Learning to Manage Design of Experiments

Although a very simple measure of performance was used with a human-in-the-loop to judge agent employability, future work may shift the focus on deployment of ML in the service of suggesting what sort of metrics might be useful, as well as determining which training runs have developed the most appropriate learner for a given deployment context. Implementing a ML agent to select the best trained agent for model deployment and determining heterogeneity ratios.

7.3 Evolve Measures of Performance

A relatively simple measure was used to grade the performance of the emergent collective that was quite simply the sum of the system's parts. Future work will focus on putting such aggregate measures in greater context and exploring scenarios that tend to exhibit, and benefit from, emergent properties.

7.4 Imitation Learning and Virtual Reality

Imitation learning allows one to harness the human brain's natural ability to process and integrate multiple (and often multi-modal) informational streams presented in complex environments and then make decisions under uncertainty (the global state of universe is not known to any given agent). Learning via imitation methods avoids the difficult task of explicitly stating the rules and their relative priority in all probable situations that an expert human is following. We believe that fully immersive, embodied experiences will enable a virtuous cycle of human-machine teaming where humans are training artificial agents and vice-versa.

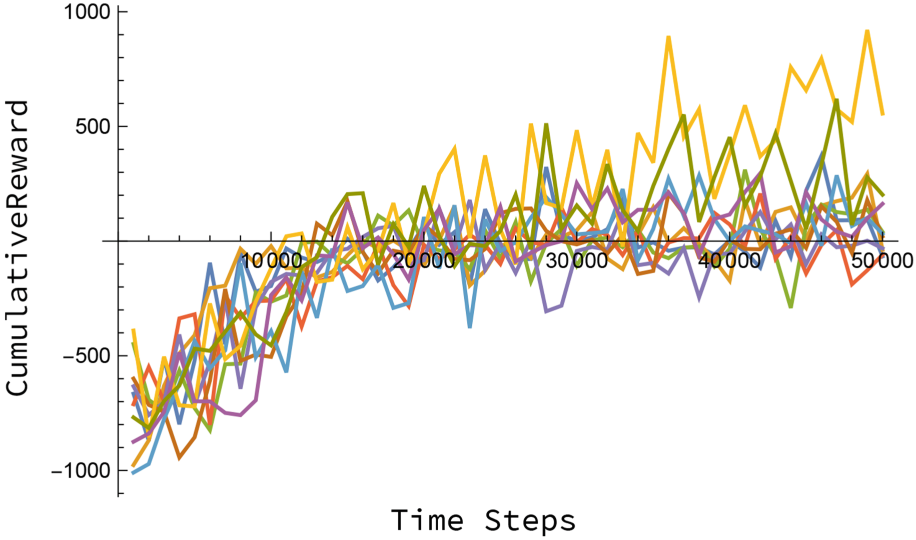


Fig. 2. Cumulative reward across 10 training runs of a stylized Search and Rescue scenario

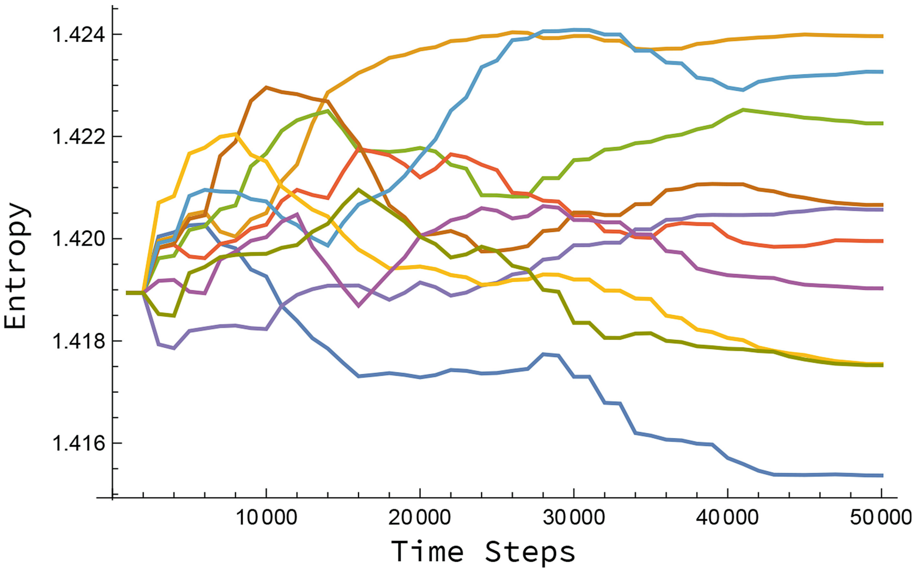


Fig. 3. Entropy across 10 training runs of a stylized Search and Rescue scenario

8 Conclusion

We have shown that agent-based modeling can be conducted using reinforcement learning agents within a fully-featured 3D game environment. Our trained agents, on the whole, were successful in accomplishing their mission. We have learned that these initial attempts at training, while showing promise, still produce a range of behavioral features and resulting cumulative rewards across runs. This indicates there is much work yet to be done to understand how the complexities of the environment, the interactions between agents, and the learning hyperparameters all come together to produce the results we have seen.

The INCOSE complexity primer [13] recommends the use of a variety of methods and approaches from a variety of disciplines that have to cope with complexity in support of analyzing, diagnosing, modeling, and synthesizing complex systems. We followed these recommendations in our research and applied machine learning, agent based models, and simulation-enabling game engines in support of architectural analysis challenges, demonstrating how the combination of these methods supports the complex systems engineering processes captured in the use case of autonomous search and rescue. This successful application shows the usefulness of this approach for a new category of engineering challenges requiring the application of complexity science.

Acknowledgements and Disclaimer. The work presented in this paper was partly supported by the MITRE Innovation Program. The authors wish to specifically acknowledge Sham Chakravorty and Muhammad Sungkar for their contributions to this effort. The views, opinions, and/or findings contained in this paper are those of The MITRE Corporation and should not be construed as an official government position, policy, or decision, unless designated by other documentation. It is approved for Public Release; Distribution Unlimited. Case Number 17-3081-17.

References

1. Cognata. http://www.cognata.com
2. Diallo, S., Mittal, S., Tolk, A.: Research agenda for next-generation complex systems engineering. In: Emergent Behavior in Complex Systems Engineering: A Modeling and Simulation Approach, pp. 379–397 (2018)
3. Epic. https://www.unrealengine.com
4. Holland, J.H., Miller, J.H.: Artificial adaptive agents in economic theory. Am. Econ. Rev. **81**(2), 365–370 (1991)
5. HTC. https://www.vive.com
6. Kelso, J.S.: Dynamic Patterns: The Self-organization of Brain and Behavior. MIT Press, Cambridge (1997)
7. Liebovitch, L.S., Peluso, P.R., Norman, M.D., Su, J., Gottman, J.M.: Mathematical model of the dynamics of psychotherapy. Cogn. Neurodyn. **5**(3), 265–275 (2011)
8. Norman, M.D.: Complex systems engineering in a federal it environment: lessons learned from traditional enterprise-scale system design and change. In: 2015 9th Annual IEEE International Systems Conference (SysCon), pp. 33–36. IEEE (2015)

9. Norman, M.D., Koehler, M.T., Pitsko, R.: Applied complexity science: enabling emergence through heuristics and simulations. In: Emergent Behavior in Complex Systems Engineering: A Modeling and Simulation Approach, pp. 201–226 (2018)
10. NVIDIA. https://www.nvidia.com
11. Reynolds, C.W.: Flocks, herds and schools: a distributed behavioral model. In: ACM SIGGRAPH Computer Graphics, vol. 21, pp. 25–34. ACM (1987)
12. Schulman, J., Wolski, F., Dhariwal, P., Radford, A., Klimov, O.: Proximal Policy Optimization Algorithms. ArXiv e-prints (2017)
13. Sheard, S., Cook, S., Honour, E., Hybertson, D., Krupa, J., McEver, J., McKinney, D., Ondrus, P., Ryan, A., Scheurer, R., et al.: A complexity primer for systems engineers. INCOSE Complex Systems Working Group White Paper (2015)
14. Silvey, P.E., Norman, M.D.: Embodied cognition and multi-agent behavioral emergence. In: Proceedings of the Ninth International Conference on Complex Systems (ICCS 2018) (2018, in press)
15. Strogatz, S.: Sync: The Emerging Science of Spontaneous Order. Penguin, London (2004)
16. Taleb, N.N.: Antifragile: Things That Gain from Disorder, vol. 3. Random House Incorporated (2012)
17. Tampuu, A., Matiisen, T., Kodelja, D., Kuzovkin, I., Korjus, K., Aru, J., Aru, J., Vicente, R.: Multiagent cooperation and competition with deep reinforcement learning. PLoS ONE **12**(4), e0172395 (2017)
18. TensorFlow. https://www.tensorflow.org
19. Tolk, A.: The next generation of modeling & simulation: integrating big data and deep learning. In: Proceedings of the Conference on Summer Computer Simulation, pp. 1–8. Society for Computer Simulation International (2015)
20. Tolk, A., Diallo, S., Mittal, S.: Complex systems engineering and the challenge of emergence. In: Emergent Behavior in Complex Systems Engineering: A Modeling and Simulation Approach, pp. 79–97 (2018)
21. Tolk, A., Rainey, L.B.: Toward a research agenda for m&s support of system of systems engineering. In: Modeling and Simulation Support for System of Systems Engineering Applications, pp. 581–592 (2015)
22. Unity3D. https://www.unity3d.com
23. Unity3D. https://github.com/Unity-Technologies/ml-agents
24. Xu, J.: Model calibration. In: Advances in Modeling and Simulation: Seminal research from 50 Years of Winter Simulation Conferences, pp. 27–46 (2017)

Forest Complexity in the Green Tonality of Satellite Images

Juan Antonio López-Rivera[1,2(✉)], Ana Leonor Rivera[1,3], and Alejandro Frank[1,3,4]

[1] Centro de Ciencias de la Complejidad, Universidad Nacional Autónoma de México, Mexico City, Mexico
jantonio.lr@ciencias.unam.mx
[2] Facultad de Ciencias, Universidad Nacional Autónoma de México, Mexico City, Mexico
[3] Instituto de Ciencias Nucleares, Universidad Nacional Autónoma de México, Mexico City, Mexico
[4] El Colegio Nacional, Mexico City, Mexico

Abstract. Forest complexity is associated with biodiversity and tells us information about the ecosystem health. A healthy forest must be in a scale-invariant state of balance between robustness and adaptability, reflected in the tonalities present on its vegetation. Remote imaging can be used to determine forest complexity based on the scale-invariance of green tones in the images. Here is proposed a simple technique to monitor changes on the forest using statistical moments and spectral analysis of the green tones on the satellite images.

Keywords: Forest complexity · Ecosystem health · Scale invariance

1 Introduction

Forests are one of the most important ecosystems for environment conservation. Specifically, forest "health" can be used to understand the damage produced by climate change on the Earth, fires, modification of land-use or unsustainable management due to anthropogenic causes [1–3]. Monitoring of forests' health is done by expensive and locally limited in-situ observations or by global remote sensing using satellite images [4, 5]. Forest development can be seen in two different images [6, 7], and through different moments by time series analysis [8, 9]. Remote sensing has been popularized due to the improvement on computation capabilities and the free open access to Landsat and other satellite data [10–13].

Our hypothesis is that health can be defined as the critical state of balance between robust and stochastic dynamics [14], and can be determined by spectral analysis as scale-invariant distributions. We have used this hypothesis on humans [15, 16], and here we proposed it for ecological systems like a forest. In particular a healthy forest must be in a scale-invariant state reflected in the tonalities present on its vegetation. The original idea in this presentation is that remote sensing can be used to determine forest complexity based on the scale-invariance of green tones in the images using statistical moments and spectral analysis. The proposed methodology is simple, fast and cheap. Here we use

A. J. Morales et al. (Eds.): ICCS 2018, SPCOM, pp. 184–188, 2018.
https://doi.org/10.1007/978-3-319-96661-8_19

Landsat images to study forest colors to determine parameters for its characterization that can help to monitor changes.

2 Methods

Landsat images of different forest are analyzed with a simple Python script that transforms each pixel´s RGB hexadecimal value into its corresponding decimal number. This gives us a matrix for which statistical moments, power spectral density, and Shannon entropy are evaluated to find characteristic parameters of the region. We analyze as many as possible images of the same region (same latitude and longitudinal coordinate rectangle, see Fig. 1) from different dates and times to determine changes in a particular forest. Seasonal variations caused by solar angle differences and vegetation changes are one of the principal causes of noise in remote sensing change detection. To avoid this, we select images from the same season of the year. Clouds and their shadows complicate the use of Landsat data, thus for detection of forest changes, they must be removed. However, there is no algorithm that can remove all types of clouds or their shadows [9], even the most sophisticated machine learning ones [17]. We circumvent this by only using images taken without clouds.

Fig. 1. Landsat image of the Zempoala forest, 2017.

3 Results and Discussion

We analyzed images from different forests on the Earth, typical examples of "healthy" forests (Gambela National Park in Ethiopia, Payette National Forest, Idaho in USA and Zempoala Lagoons in Mexico). We also evaluated images of the Canary Islands before and after a big forest fire occurred in the summer of 2008 (see Fig. 2). Images covered from 2002 to 2017, all correspond to the January of each year. Statistical results, and slope of the power spectral density (PSD) curve are reported on Table 1.

In Fig. 2 is clear the difference on green tonality is clear years before and after the forest fire, the histogram changes from a Gaussian distribution to a Poissonian one, with almost white noise behavior in the spectral domain. In Table 1, it is notable that the images of Canary Island after the forest fire have similar statistics than "healthy" forests.

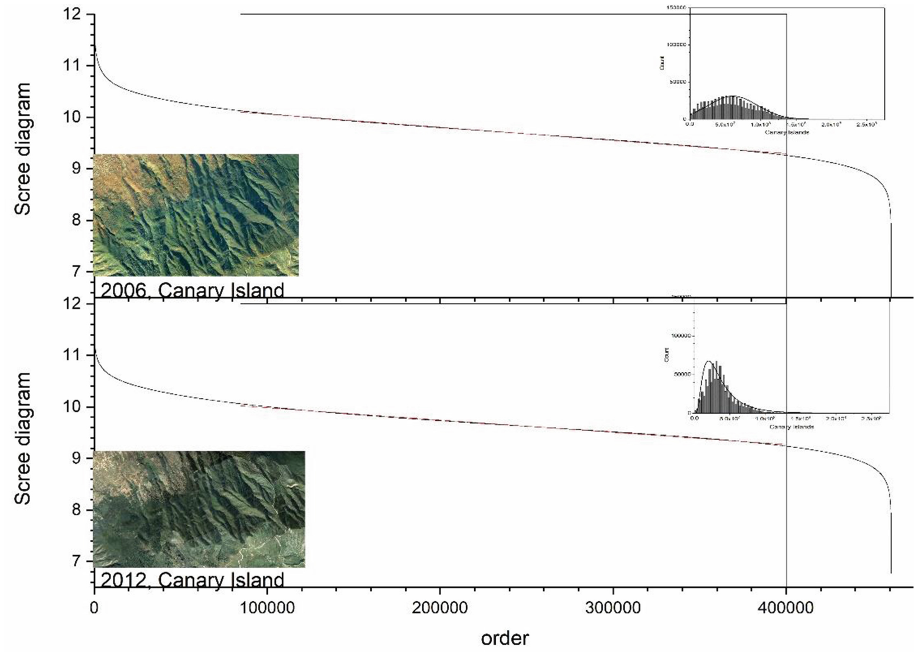


Fig. 2. Scree diagram of the Landsat images of Canary Island before (upper panels) and after (bottom panels) the big forest fire of 2008. Insert are the images (left) and the histogram (right). Histogram change to more rigid distribution, and slope decreases, indications of criticality lose.

Table 1. Green scale data of Landsat forest image. Reported values are average ± standard deviation of the statistical moments and the slope of the scree diagram obtained from the spectral analysis.

Region	Coefficient of variation	Skewness	Kurtosis	Slope
Canary Island (before forest fire)	49 ± 9	0.5 ± 0.4	0.3 ± 0.8	−3E-06
Canary Island (after forest fire)	46 ± 9	1.8 ± 0.4	9 ± 3	−2E-06
Zempoala lagoons, Mexico	47 ± 19	1.9 ± 0.3	7 ± 4	−2E-06
Payette National Forest, Idaho, USA	42 ± 12	3.2 ± 0.9	25 ± 13	−2E-06
Gambela National Park, Ethiopia	34 ± 6	12 ± 2	197 ± 67	−2E-06

4 Conclusion

In this project, we propose a methodology to determine the forest “health” using satellite images. Statistical values of the green tones present on the images are significantly different for the same region in images taken years before and after a forest fire. Moreover, the distribution changes from Gaussian (almost symmetric, mesokurtic) to a Poissonian one (asymmetric with long tails, leptokurtic) indicating change from an independent variable to a stochastic one. Moreover, other forests that can be considered “un-healthy” have clearly Poissonian distributions, with a characteristic white noise signal in the spectral domain. This technique is promising and needs to be explored in different forests and instants. So far results indicate this approach can be useful as a first glance of a forest’s state.

Acknowledgments. Partial Financial support for this work was provided by UNAM through grants DGAPA-PAPIIT IN106215, IV100116, and by CONACyT Frontera grants FC-2016-1/2277.

References

1. Laurance, W.F., Camargo, J.L., Fearnside, P.M., Lovejoy, T.E., Williamson, G.B., Mesquita, R.C., et al.: An Amazonian rainforest and its fragments as a laboratory of global change. Biol. Rev. **93**(1), 223–247 (2018). https://doi.org/10.1111/brv.12343
2. Seidl, R., Thom, D., Kautz, M., Martin-Benito, D., Peltoniemi, M., Vacchiano, G., et al.: Forest disturbances under climate change. Nat. Clim. Change **7**(6), 395 (2017). https://doi.org/10.1038/nclimate3303
3. Reyer, C.P., Brouwers, N., Rammig, A., Brook, B.W., Epila, J., Grant, R.F., et al.: Forest resilience and tipping points at different spatio-temporal scales: approaches and challenges. J. Ecol. **103**(1), 5–15 (2015). https://doi.org/10.1111/1365-2745.12337
4. Lausch, A., Erasmi, S., King, D.J., Magdon, P., Heurich, M.: Understanding forest health with remote sensing-Part II—A review of approaches and data models. Remote Sens. **9**, 129 (2017). https://doi.org/10.3390/rs9020129
5. Banskota, A., Kayastha, N., Falkowski, M.J., Wulder, M.A., Froese, R.E., White, J.C.: Forest monitoring using landsat time series data: a review. Can. J. Remote Sens. **40**, 362–384 (2014). https://doi.org/10.1080/07038992.2014.987376
6. Singh, A.: Digital change detection techniques using remotely-sensed data. Int. J. Remote Sens. **10**, 989–1003 (1989). https://doi.org/10.1080/01431168908903939
7. Coppin, P., Jonckheere, I., Nackaerts, K., Muys, B., Lambin, E.: Digital change detection methods in ecosystem monitoring: a review. Int. J. Remote Sens. **25**, 1565–1596 (2004). https://doi.org/10.1080/0143116031000101675
8. Kennedy, R.E., Andrefouet, S., Cohen, W.B., Gomez, C., Griffiths, P., Hais, M., et al.: Bringing an ecological view of change to Landsat-based remote sensing. Front. Ecol. Environ. **12**, 339–346 (2014). https://doi.org/10.1890/130066
9. Zhu, Z.: Change detection using Landsat time series: a review of frequencies, pre-processing, algorithms, and applications. ISPRS J. Photogram. Remote Sens. **130**, 370–384 (2017). https://doi.org/10.1016/j.isprsjprs.2017.06.013
10. Woodcock, C.: Free access to Landsat imagery teach by the book science education. Science **320**, 1011–1012 (2008). https://doi.org/10.1126/science.320.5879.1011a

11. Wulder, M.A., White, J.C., Loveland, T.R., Woodcock, C.E., Belward, A.S., Cohen, W.B., et al.: The global Landsat archive: status, consolidation, and direction. Remote Sens. Environ. **185**, 271–283 (2016). https://doi.org/10.1016/j.rse.2015.11.032
12. White, J.C., Wulder, M.A., Vastaranta, M., Coops, N.C., Pitt, D., Woods, M.: The utility of image-based point clouds for forest inventory: a comparison with airborne laser scanning. Forests **4**, 518–536 (2013). https://doi.org/10.3390/f4030518
13. Hartig, T., Mitchell, R., De Vries, S., Frumkin, H.: Nature and health. Ann. Rev. Public Health **35**, 207–228 (2014). https://doi.org/10.1146/annurev-publhealth-032013-182443
14. Rivera, A.L., Estañol, B., Toledo, J.C., Fossion, R., Frank, A.: Looking for biomarkers in physiological time series. In: Quiroz, L.O., Antonio, O.R. (eds.) Quantitative Models for Microscopic to Macroscopic Biological Macromolecules and Tissues. Springer, Cham (2018)
15. Rivera, A.L., Estañol, B., Fossion, R., Toledo-Roy, J.C., Callejas-Rojas, R.C., Gien, J.A., et al.: Loss of breathing modulation of heart rate variability in patients with recent and long-standing Diabetes Mellitus Type II. PLoS ONE **11**(11), e0165904 (2016). https://doi.org/10.1371/journal.pone.0165904
16. Rivera, A.L., Estañol, B., Sentíes-Madrid, H., Fossion, R., Toledo-Roy, J.C., Mendoza-Temis, J., et al.: Heart rate and systolic blood pressure variability in the time domain in patients with recent and long-standing Diabetes Mellitus. PLoS ONE **11**(2), e0148378 (2016). https://doi.org/10.1371/journal.pone.0148378
17. Hughes, M.J., Hayes, D.J.: Automated detection of cloud and cloud shadow in single-date Landsat imagery using neural networks and spatial post-processing. Remote Sens. **6**, 4907–4926 (2014). https://doi.org/10.3390/rs6064907

Embodied Cognition and Multi-Agent Behavioral Emergence

Paul E. Silvey(✉) and Michael D. Norman

The MITRE Corporation, Bedford, MA 01730, USA
{psilvey,mnorman}@mitre.org

Abstract. Autonomous systems embedded in our physical world need real-world interaction in order to function, but they also depend on it as a means to learn. This is the essence of artificial Embodied Cognition, in which machine intelligence is tightly coupled to sensors and effectors and where learning happens from continually experiencing the dynamic world as time-series data, received and processed from a situated and contextually-relative perspective. From this stream, our engineered agents must perceptually discriminate, deal with noise and uncertainty, recognize the causal influence of their actions (sometimes with significant and variable temporal lag), pursue multiple and changing goals that are often incompatible with each other, and make decisions under time pressure. To further complicate matters, unpredictability caused by the actions of other adaptive agents makes this experiential data stochastic and statistically non-stationary. Reinforcement Learning approaches to these problems often oversimplify many of these aspects, e.g., by assuming stationarity, collapsing multiple goals into a single reward signal, using repetitive discrete training episodes, or removing real-time requirements. Because we are interested in developing dependable and trustworthy autonomy, we have been studying these problems by retaining all these inherent complexities and only simplifying the agent's environmental bandwidth requirements. The Multi-Agent Research Basic Learning Environment (MARBLE) is a computational framework for studying the nuances of cooperative, competitive, and adversarial learning, where emergent behaviors can be better understood through carefully controlled experiments. In particular, we are using it to evaluate a novel reinforcement learning long-term memory data structure based on probabilistic suffix trees. Here, we describe this research methodology, and report on the results of some early experiments.

Keywords: Embodied cognition · Reinforcement learning
Agent-based modeling · Multi-agent systems · Emergence

1 Introduction

Autonomous systems embedded in our physical world need to be intellectually competent to perform the tasks they were designed for, but need not demonstrate

A. J. Morales et al. (Eds.): ICCS 2018, SPCOM, pp. 189–201, 2018.
https://doi.org/10.1007/978-3-319-96661-8_20

human-level abilities with language or reasoning to be effective and valuable. They will, however, need to function through real-world interaction and, most likely, will gain the knowledge on which they depend from direct experiential learning. Artificial agents facing adversarial challenges need cognitive capabilities that are tightly integrated with, and dependent on, their embodiment, but today's dominant Machine Learning (ML) methods have limited applicability in these kinds of dynamic situated environments, which are characterized as continuous, on-going, time-pressured, uncertain, and statistically non-stationary [15,28]. In response, we are developing and testing holistic, temporally sensitive machine learning and goal-directed planning methods for embodied cognition and multi-agent Artificial Intelligence (AI), with the objective of improving our ability to build resilient and trustworthy real-world autonomous systems. To this end, we have built a Java testbed designed to host a variety of progressively more difficult challenge scenarios, in which multiple artificial agents play cooperative, competitive, and adversarial games with each other and their environments. While this is primarily intended to offer a means to test and refine our episodic and procedural memory innovations, it naturally presents a dual opportunity to study behavioral emergence and complexity in a bottom-up fashion [24].

The best examples we have of intelligent behavior are biological ones, particularly animate beings with long lifespans of embodied learning in our physical world. Insects, animals, and humans all have multi-modal sensory organs coupled to neural information processing systems, as well as real-world effectors that allow them to both change their perspectives and to alter their local environments. Proponents of Artificial General Intelligence argue that AI has become too compartmentalized and fragmented, where researchers study things like vision, language, reasoning, and planning in isolation. Even Machine Learning is fragmented into significantly diverse sub-schools of technique and thought [11]. Primitive embodied cognition, easily observable in even the lowest life forms, requires that many of these faculties be integrated into what we might call a system of systems. One theme that stands out for such embodied cognitive agents is the importance of time. Data comes to their sensory mechanisms as impressions or signals that vary with time, and the perceived world can change at varying rates, often completely independent of the agent's presence. By directly confronting the temporal aspects of situated learning and decision making, and by integrating these capabilities, we believe we can influence and improve the body of practice in building trustworthy and autonomous systems.

Our bottom-up methodology is motivated by drawing a behavioral research analogy with the Study of Model Organisms in the Life Sciences, where many groundbreaking discoveries have been made. Here, cellular and genomic research is more easily and effectively conducted using simple life forms, such as the fruit fly or nematode [17]. Similarly, our artificial agents are simple enough for us to conduct extensive behavioral and learning experiments with them, yet they are embodied in realistically complex worlds they can only partially observe, understand, and control. Coping with limited sensory, memory, and processing time resources is a hallmark of intelligence. Our approach enables us to study and to test learning under such real-world cognitive constraints. We specifically start with low-entropy, low-bandwidth sensory data to allow rapid experimentation using standard compute resources. Our agents use relatively simple data-driven mechanisms to construct problem-specific and accurate Markov Decision Process models and use behavioral learning algorithms to cope with changing temporal sequence data and real-time dynamic pressure.

This paper is organized as follows: first, we review some significant past AI research in embodied cognition and reinforcement learning and discuss emergent behaviors in multi-agent simulations. Then, our own research testbed is described, along with experimental scenarios that we have explored. Next, our holistic cognitive architecture is introduced, and we then share some of the unexpected learned behaviors our agents presented for us to analyze. Finally, we conclude with our plans to continue this work.

2 Related Work

The disciplined Computer Science quest to develop true artificial intelligence is over sixty years old, but during most of that time was dominated by disembodied mind research, where a focus on language and knowledge in many ways led researchers to detach aspects of thinking and reasoning from sensing, acting, and real-world interaction. However, there have always been those who believed intelligence depends in non-trivial ways on being embodied and learning from the uncertain experiences of living.

For example, by the late 1980s, seminal work at MIT had begun to build simple insect-like robots [7], which served to demonstrate the power of primitive reactive thinking mechanisms that could work without deliberative or goal-directed planning or learning. Soon there was great interest in the idea of autonomous agents and distributed AI, often using object-oriented programming concepts to encapsulate and situate artificial agents in simulated or virtual problem environments [13]. A good summary of the development of these views can be found in Anderson's field guide to embodied cognition [2]. At the same time, clearer definitions and better algorithms for Reinforcement Learning as a distinct paradigm were being developed [25], while others were beginning to pursue universal and general cognitive architectures [14].

Machine learning is now commonly seen to include separate paradigms for supervised, unsupervised, and reinforcement learning. Deep learning has mostly been applied in supervised learning contexts, but has increasingly been used as a way to support reinforcement learning in games with enormous state spaces [23]. Closer in spirit to our approach is the combination of temporal sequence memory-based pattern recognizers, minimum description length modeling, and reinforcement learning [18].

Another area of inspiration for us, where simplicity has shown great power, is complexity science. Here, chaos can be generated by something as simple as the recurrence equations of the logistic map, and patterns that are unpredictably complex can be generated by simple one-dimensional cellular automata [16].

Likewise, the study of the emergence of cooperation, which we discuss more later in this paper, came from simple game-theoretic and simulation analysis [3]. It has now become standard practice to try to understand complex systems using Agent-Based Modeling.

3 The MARBLE Framework and Testbed

Our Multi-Agent Research Basic Learning Environment (MARBLE) is a discrete modeling framework and simulation testbed for doing embodied cognition research, implemented in the Java programming language. It is constructed around the Model-View-Controller design pattern [9], based on a 2D world model consisting of a hexagonal grid that, by default, wraps around on all six sides. The wrap-around logic is such that moving forward in any one direction repeatedly will visit each cell in the world exactly once before returning to the starting cell [21], but these are otherwise arbitrary choices for a partially observable environment with spatial and temporal characteristics. Agents and other simulated objects may occupy any given cell, and agents have directional orientation with a limited field of view and a small repertoire of possible actions on any given simulation time step. We use a reinforcement learning paradigm, where each agent receives sensory data and reward signals for the results of their prior-step action choice, and all agents attempt their next moves simultaneously. This forces the environment controller to mediate the results of agent actions, such that the agents must discover, through experience, what their actions might or might not accomplish for them in particular discernible situations. Agents receive their sensory data as a *DataFrame* object, which is a bit-string of problem-specific length whose structure remains undisclosed to the agents, encoding experimenter-determined discernible characteristics or perceptual features within the agent's field of view. For example, a simple world might use only 3 bits to encode whether each of the three immediately facing cells contains an object or not. Slightly more information-rich examples might use 3 bits per cell for a total *DataFrame* size of 9 bits, where each cell can distinguish eight different states, including being empty, containing a food object, or containing another agent in one of its six directional orientations. See Fig. 1 for some currently established MARBLE elements.

We have implemented two problem scenarios, both related to a simple food gathering learning challenge. The first contains only a single agent, who lives in a world that contains only one other object, which is a green cell representing food. The agent can choose between one of the following four distinct actions per time step: stepping forward one cell, turning 60° right or left, or doing nothing. The reward structure for this game is −1 point for any action that does not result in the agent stepping into the cell containing food, and +1000 points when it consumes the food in this way. When the food is consumed, a new food object immediately appears four cells directly ahead of the agent, based on its orientation at the time. Optimal behavior for this problem is easy to specify, unless the size of the hex grid becomes small enough that the shortest sequence

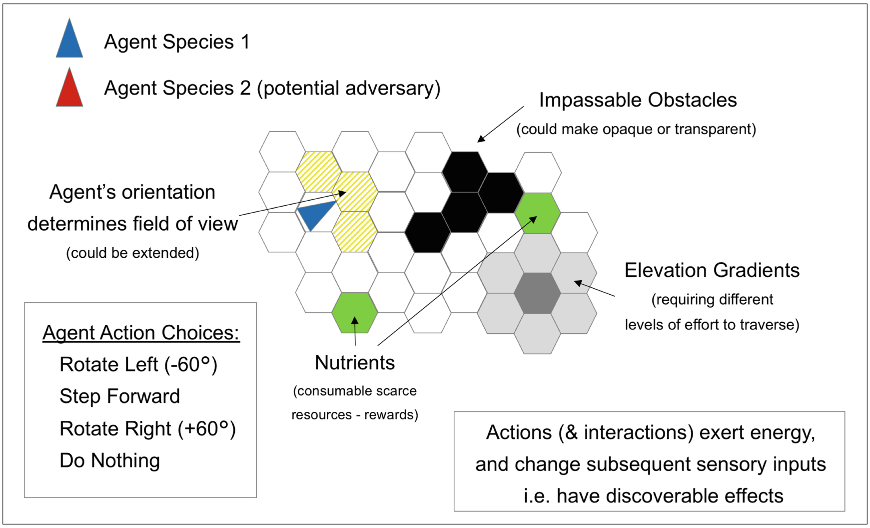


Fig. 1. Example MARBLE elements.

of moves between feedings is not simply to keep walking forward. This problem is completely deterministic, and only poses a simple delayed-gratification challenge. In this case, the sensory data presented to the agent at each time step is the 3-bit example given above, where an unoccupied cell is perceived as a zero bit, and the existence of the food object in any of the three cells facing the agent (immediately ahead of and to its left and right periphery) is perceived as a one bit (see agent's field of view in Fig. 1).

The second variant of the food gathering problem consists of a red agent and a blue agent who co-exist in the same otherwise empty world, also with a single food object that they must compete to find. The reward structure is similar, with -1 point for staying put, turning, or stepping into an empty cell, and $+1000$ points for consuming the food. The food regenerates as before, four steps ahead of the agent who has just consumed it, or one additional step forward if that cell is occupied by the other agent. Our physics rules for this problem prevent the two agents from occupying the same cell at the same time, so conflict resolution logic must be added to the controller. When both agents attempt to step into the same cell at the same time, a random coin flip determines who will succeed, with the losing agent not moving from its cell of origin and receiving a -10 point reward. The winner will receive either the usual -1 point for entering an empty cell, or $+1000$ points if the cell contains the food object. This problem was intended to study competition, and as such the agents are penalized -50 points for the aggressive action of trying to move into a cell that the other agent already occupies (and is not simultaneously vacating). If the two agents are facing one another and both try to move forward, they each receive the -50-point penalty. The existence of another competitive agent changes the stochastic properties of the world considerably, and, as we will see, creates the opportunity for many complex behaviors to emerge. The sensory data for this problem is also more feature-rich, with each agent being able to discriminate empty cells, cells occupied by the food object, and cells occupied by the other agent (as well as the other agent's relative orientation). This uses a 9-bit *DataFrame* as described earlier.

4 Embodied Agent Learning

Our MARBLE agents are built using a long-term episodic and procedural memory data structure which records temporal sequence traces, maintaining event distribution statistics for observed successor states and expected discounted rewards for actions that have been tried. Our ML research is aimed at understanding how well these memory structures perform in a variety of challenge problem environments, and how well certain parameter settings might affect rates of learning, abilities to generalize effectively, and abilities to cope with non-stationarities caused by the presence of, and interactions with, other adaptive agents.

The basis of our learning framework is a variable depth probabilistic suffix tree [26], which produces an environmentally data-driven variable-order Markov Model [5], using focus bits from each data frame in the temporal sequence sensory data. For large data frames, multiple trees using different foci bits can be built in parallel, and their independently recommended actions can be pooled using an ensemble strategy. In low-entropy environments or when the number of focus bits is small, these trees are sparse and can grow deep to discover causal patterns with long temporal lags. Each node in the tree represents a state with probabilistic knowledge of its successors, updated through experience in a Bayesian manner[1]. As in hidden Markov models, these tree nodes have both output variable probabilities for expected rewards, as well as successor (hidden) state transitions within the tree. These internal transitions are computed (vs. explicitly being stored) by walking from a current state leaf node to the root, appending the latest perceptual state as a new suffix to a short term memory buffer, and using that new suffix to walk down the tree to the next leaf node state. The current-to-next leaf node pairs represent logical links within the tree that, when considered separately, form a partial DeBruijn Graph, since the short-term memory buffer acts like a shift register as time progresses.

Two types of learning are implemented in our MARBLE agents, and they both occur by updating probability estimates while ascending from the current leaf node to the tree root as just described. The first can be described from the perspective of model-free design [19], where the expected utility of the most recent action is updated using the Q-Learning technique [27]. It is model-free in the sense that it blindly learns actions that will maximize rewards without any attempt to develop sensory expectations. Learning is completely based on

[1] Posterior probabilities are computed from maximum entropy priors initialized by setting the alpha parameter in a multi-modal Dirichlet distribution.

event sequences the agent has actually experienced. The other type of learning is model-based, as it updates the probability of seeing the just-experienced perceptual symbol extracted from the *DataFrame* for each successively shorter suffix state (defined by the tree nodes visited during the climb to the root).

This second form of learning prepares the agent to use its memory in imaginative ways and to envision futures that are probabilistically plausible, whether they in fact have ever actually been experienced. Probabilistic suffix trees have been used in compression algorithms because they construct statistical prediction models (which can inform arithmetic or Huffman encoders based on a symbol's probability of occurrence). This type of learning forms the basis of deliberative planning mechanisms, including ones like Monte Carlo planning [8], which can be implemented to work in time-constrained situations as an Anytime Algorithm [10]. Neuroscientists and AI researchers have hypothesized that temporal sequence prediction is a core foundational capability of the brain's neocortex, common across all sensory data modalities [12]. Similarly, the ability to recognize surprise in the form of expectation failures has been seen as a significant trigger mechanism for attentionally-directed learning of event salience [22]. As a result, this kind of model-based learning is seen as extremely important for inclusion in embodied cognitive agent research.

Figure 2 shows the global view of the MARBLE competitive problem scenario, with the blue agent highlighted and the memory load/store menu item selected. The probabilistic suffix tree memories can thus be saved to disk and reloaded for additional training and/or testing at a later time. The controller buttons at the top include a single step button and a run/stop button with radio controls for two speeds. Clicking on an agent when the simulation is stopped highlights it with a yellow outline and allows the user to 'drive' that agent manually using the arrow cursor keys (which also invoke the step operation). In addition, a right-click on an agent will bring up a parameter dialog box where learning modes and exploration rates and conditions can be set (see Fig. 2 dialog box).

As noted, cognitive constraints include both limitations to memory space and to thinking time. They are naturally related in MARBLE agents because algorithms that learn through traversal of the memory tree, or creatively plan by using it to generate hypothetical futures, will take more time as the tree grows. Although there are many relatively low-entropy problem domains that an embodied cognitive agent might face, distinguishing all observable state combinations over long time horizons without constraint produces unbearable combinatoric growth. To address this, our research plan includes several methods we want to test to implement controlled remembering, selective forgetting, and generalized abstraction via state combination, the details of which are beyond the scope of the current paper. As our initial experiments are with limited bandwidth and small-world environments, this potential scaling problem has not yet arisen.

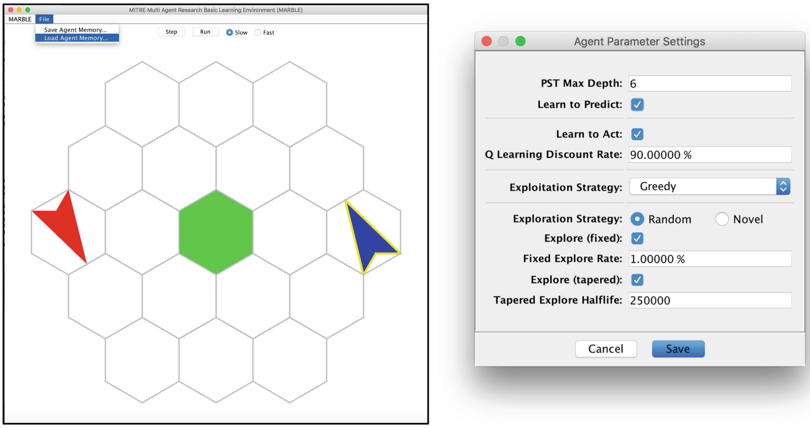


Fig. 2. Small-world MARBLE competition with agent parameters.

5 Behavioral Emergence

In this section, we describe three examples of emergent behavior witnessed from our experiments. We use colloquial descriptions of human-like behavior to illustrate how we interpreted what the agents were doing.

Our first example came from asking how best to determine what had been learned. Instead of simply looking at metrics like total accumulated reward or average reward per step (which should improve over time), we wanted to test the agent from situations where we knew what we would have done, by thinking of the food not just as a perceptual bit pattern but as an object in the world. We observed that agents who explored a lot appeared to learn the concept Psychologists refer to as Object Permanence (with respect to the food, which persists in its cell until consumed in the single agent scenario), whereas those who always tried to exploit their acquired knowledge clearly did not. We could see this by manually controlling the agents to put them into test states, from which we let them decide how to act. For example, if the food was in their right periphery but we turned them left (away from the food), smarter agents knew that two successive right turns would return them to facing it, demonstrating their grasp of Object Permanence. Similarly, we tested a number of drive-by or near-miss food encounters, where memory of the recent past was exploited by the most intelligent agents, while others seemed to possess ineffectual or missing short-term memories. Here, two factors were relevant for the smarter agents. First, as noted, they explored more, and second, they had suffix tree memories that were deep enough to distinguish and remember what they had observed multiple time-steps in the immediate past. Cognitive resources such as memory capacity are necessary but not sufficient for intelligence, and they attain their ultimate value for agents by the diversity of experiences that feed them.

Our second example is one of stubbornness. Our competitive learning scenario with two agents was designed to present a non-stationary learning problem that was as simple as possible. The initial learning algorithms we tested, however, were not designed for this, since they assumed that estimated probabilities would converge by the law of large numbers. Our training cases usually pitted novice agents against each other, often with the exact same mental resources and parameter settings. The rate of exploration (vs. exploitation) was slowly tapered to vanishing, and the learned greedy mode behaviors were examined at the end of the run. As is common in many machine learning situations, we witnessed numerous cases where behaviors appeared to have settled into a local (vs. global) optimum. When exploration stopped, it became apparent that some of these behaviors led to starvation for one or both agents, even though they both survived reasonably well whenever there were still some occasional random actions taken. The most unexpected of these happened with two agents that had shallow suffix trees of only depth 2 and when they both learned that going forward was the best action in almost all cases. Because their expected rewards had converged through many trials, if they ran into each other they got locked into a head-butting stalemate, which only could be broken by many steps of receiving the −50-point penalty. We believe that faster forgetting, in the form of non-stationary reward estimators, will avoid this behavior or at least permit it to be seen as futile more quickly.

Finally, in several competitive agent experiments conducted using the small 19-cell world, the agents essentially learned to cooperate, by dancing a kind of waltz (there were three time-steps between feedings). We had hypothesized that some synergistic or win-win behavior might emerge, as opposed to the pathologies just described, but this behavior appeared remarkably clever. The agents alternated in consuming the food, carefully positioning themselves so as to do this most effectively and using the wrap-around nature of the world to continue this way indefinitely. Furthermore, this behavior had become robust in the sense that random perturbations to one or the other agent upset the dance for a few steps, but they were able to settle back into the rhythm fairly soon.

6 Future Directions

Our MARBLE testbed is designed to support problems that involve many agents, with many more discernible conditions, objects, and interaction rules. We have only begun to scratch the surface of experimental conditions we can configure and test. A couple of important areas only mentioned in passing that we would like to address soon include the following.

First, we need to test the variety of methods we have planned for dealing with non-stationarities, including fading memories using moving window averages or exponentially decaying weighted ones. In addition, there are numerous techniques that data analysis practitioners have developed for measuring the degree of stationarity in time-series data, which could likely be employed as adaptive hyper-parameter setting mechanisms.

Multi-goal (or multi-task) learning is a real-world problem and a topic that is currently of great interest in machine learning research. It is more complicated than simply training a single classifier to recognize multiple types of objects, or a single neural network to play a variety of games without retraining, though these problems pose significant challenges in their own right. Recognizing that motivations are neither constant nor linear (e.g., consumption affects appetite), and that goals can interact in both synergistic as well as antagonistic ways, are just a few concerns. Sometimes goals can be stacked in a kind of hierarchy, with lower level or immediate need (tactical) goals taking precedence over longer-range (strategic) goals. Making curiosity a motivation to drive exploration, after more primitive needs such as safety and survival have been met, is one example of this type of thinking. On the other hand, it is interesting to speculate how many people would survive their childhood if they didn't have adults overseeing their play, suggesting this problem might best be solved in a social or multi-agent manner. See [4] for a cognitive architecture design that seeks to address motivation as a first-class problem. Receiving multi-goal-indexed reward signals and maintaining separate expected utilities for them can possibly provide a foundation for flexible multi-task behaviors, and a meta-observation process might be able to learn when goals align or diverge in their recommended actions. These are some of the ideas we would like to pursue.

There have been several efforts to create standardized challenge problems for AI researchers to compare their algorithms and implementations, and we are considering an Open Source release of MARBLE for this purpose. At the same time, adapting our agents to other reinforcement learning and game simulation environments is of interest to us. Two of these are the OpenAI Gym environment [6] and the micro-Real-Time-Strategy (μRTS) game platform and competition [20]. Most of the Gym problems are cast as single agent vs. the world learning problems, and true multi-player games are only just emerging in that environment. On the other hand, the RTS-type games are inherently adversarial problems, but many competitors use heuristic strategies that are not necessarily improved through machine learning. We would like to see these testbeds used to examine agents learning against other agents (who are themselves simultaneously learning) in order to create the kinds of non-stationary challenges we seek to study.

Embodiment also presents a natural opportunity for comparing our artificial agent's cognitive abilities against those of humans and for exploring ways in which the agents might be taught faster through demonstration via remote human control using Virtual Reality (VR). In a complementary way, a skilled artificial agent could advise a human doing a similar task using Augmented Reality (AR). We plan to explore these ideas by building interfaces between our testbed and the popular Unity game engine [1]. This presents unique and powerful opportunities for bidirectional teaching and learning in human-computer teaming scenarios. If people can "get in the robot's head" to experience its sensory world and control its motor responses, they might guide its learning more quickly and efficaciously on how to deal with important problem situations.

Similarly, an artificial cognitive agent who has learned effective ways to solve a problem or get out of trouble could give hints and decision suggestions to humans doing the same task, whether using heads-up displays or other types of Augmented Reality interfaces.

7 Conclusions

While there has been a recent resurgence of interest in AI and Machine Learning, there remains a kind of compartmentalization where only narrow aspects of intelligence receive careful study. Supervised machine learning, as widely practiced today, focuses on building classifier systems from expert-curated labeled training data, based on assumptions of statistical stationarity and using off-line processing. On the other hand, research in Reinforcement Learning studies online learning that is at least episodic if not continuous, confronting head-on the problems of sensing and goal-directed action in the same manner as embodied cognitive agents in the real world. Modern robots (including industrial manufacturing robots, robo-soccer players, self-driving cars, etc.) can be taught to perform specific functions in reasonably controlled environments, but, once taught, they usually operate in exploitative or performance-only modes.

Learning to act (to exploit) requires (exploratory) acting to learn, and this is only one of many examples where real agents must use attentional focus or goal prioritization to simultaneously balance their potentially oppositional objectives. Model-free RL such as Q-Learning can learn to optimize behavior around a single reward signal, but because it builds no explicit model of its environment, is unable to realize when something very unexpected happens. Very little autonomous systems research has tried to integrate a full complement of real-world requirements for embodied cognition. These requirements include learning perceptual discriminators, developing sequential expectations from experience, using delayed rewards to hone reactive skills, dealing with variable causal time lags and real-time decision requirements, using multi-goal attention balancing and planning (when possible), and using surprise as a motivator to adapt and relearn. We believe all these cognitive challenges can be studied together using simple discrete simulations with few observable states and few selectable actions. The research we have described here takes this approach.

We conclude with one final anecdote that came from enacting this work. It is easy for us to anthropomorphize our simple artificial intelligent agents, because we have built them in our own image, so to speak. They too are embodied cognitive beings who experience and interact with a reasonably predictable world through a stream of temporal sequence data. The environments we create for them have been sometimes compared to god games because there are privileged views of the world and its rules that we understand (and program) which are not known by the agents. However, we sometimes see unintended consequences of our actions when we run the agents, as happened in the following case. Our initial implementation of the competitive controller included the conflict resolution rules described earlier, but accidentally failed to properly consider what to do in one rather unlikely case. This occurred when the food was being consumed by one agent and the location for its regeneration was in the path of the other agent. In the process of moving the food to a new cell, the controller put it back in the world before the second agent's move had been completed. When the second agent had already decided to step forward and the food was being relocated to that destination cell, the controller incorrectly credited both agents with the 1000-point reward on the same time-step, effectively moving the food twice in one step. This bug in our code violated the intended physics rules and was only discovered when we observed the agents collecting more reward than was 'possible'. The agents had found a 'glitch in the matrix' and exploited it, thus humbling their creators.

Acknowledgements and Disclaimer. The authors wish to thank Jason F. Kutarnia and Brittany A. Tracy for their assistance with this research. Approved for Public Release; Distribution Unlimited. Case Number 18-1473.

References

1. Unity 3d game engine. https://unity3d.com/public-relations
2. Anderson, M.L.: Embodied cognition: a field guide. Artif. Intell. **149**(1), 91–130 (2003)
3. Axelrod, R., Hamilton, W.D.: The evolution of cooperation. Science **211**(4489), 1390–1396 (1981)
4. Bach, J.: Principles of Synthetic Intelligence PSI: An Architecture of Motivated Cognition, vol. 4. Oxford University Press, Oxford (2009)
5. Begleiter, R., El-Yaniv, R., Yona, G.: On prediction using variable order Markov models. J. Artif. Intell. Res. **22**, 385–421 (2004)
6. Brockman, G., Cheung, V., Pettersson, L., Schneider, J., Schulman, J., Tang, J., Zaremba, W.: OpenAI Gym (2016). http://arxiv.org/abs/1606.01540v1
7. Brooks, R.: A robust layered control system for a mobile robot. IEEE J. Robot. Autom. **2**(1), 14–23 (1986)
8. Chung, M., Buro, M., Schaeffer, J.: Monte Carlo planning in RTS games. In: Proceedings of IEEE 2005 Symposium on Computational Intelligence and Games, pp. 117–125 (2005)
9. Coad, P.: Object-oriented patterns. Commun. ACM **35**(9), 152–159 (1992)
10. Dean, T.L., Boddy, M.S.: An analysis of time-dependent planning. In: Proceedings of the Seventh AAAI National Conference on Artificial Intelligence, vol. 88, pp. 49–54. AAAI Press, Saint Paul (1988)
11. Domingos, P.: The Master Algorithm: How the Quest for the Ultimate Learning Machine Will Remake Our World. Basic Books, New York (2015)
12. Hawkins, J., Blakeslee, S.: On Intelligence: How a New Understanding of the Brain Will Lead to the Creation of Truly Intelligent Machines. Macmillan, London (2007)
13. Jennings, N.R., Sycara, K., Wooldridge, M.: A roadmap of agent research and development. Auton. Agent. Multi Agent Syst. **1**(1), 7–38 (1998)
14. Laird, J.E., Newell, A., Rosenbloom, P.S.: Soar: an architecture for general intelligence. Artif. Intell. **33**(1), 1–64 (1987)
15. Machado, M.C., Bellemare, M.G., Talvitie, E., Veness, J., Hausknecht, M., Bowling, M.: Revisiting the arcade learning environment: evaluation protocols and open problems for general agents. J. Artif. Intell. Res. **61**, 523–562 (2018)

16. Mitchell, M.: Complexity: A Guided Tour. Oxford University Press, Oxford (2009)
17. Mukherjee, S.: The Gene: An Intimate History. Simon and Schuster, New York (2017)
18. Nguyen, P., Sunehag, P., Hutter, M.: Context tree maximizing reinforcement learning. In: Proceedings of the 26th AAAI Conference on Artificial Intelligence. Association for the Advancement of Artificial Intelligence (2012)
19. Norman, M.D., Koehler, M.T., Pitsko, R.: Applied complexity science: enabling emergence through heuristics and simulations. In: Mittal, S., Diallo, S., Tolk, A. (eds.) Emergent Behavior in Complex Systems Engineering: A Modeling and Simulation Approach, pp. 201–226. Wiley, Hoboken (2018)
20. Ontañón, S., Barriga, N.A., Silva, C.R., Moraes, R.O., Lelis, L.H.: The first microRTS artificial intelligence competition. AI Mag. **39**(1), 75–83 (2018)
21. Patel, A.: Red blob games, hexagonal grid reference. https://www.redblobgames.com/grids/hexagons/
22. Schank, R.C.: Dynamic Memory Revisited. Cambridge University Press, New York (1999)
23. Silver, D., Schrittwieser, J., Simonyan, K., Antonoglou, I., Huang, A., Guez, A., Hubert, T., Baker, L., Lai, M., Bolton, A., Chen, Y., Lillicrap, T., Hui, F., Sifre, L., van den Driessche, G., Graepel, T., Hassabis, D.: Mastering the game of Go without human knowledge. Nature **550**(7676), 354–371 (2017)
24. Silvey, P.E.: Leveling up: strategies to achieve integrated cognitive architectures. In: Fall Symposium Series - A Standard Model of Mind: AAAI Technical Report FS-17-05, AAAI 2017, pp. 460–465 (2017)
25. Sutton, R.S.: Learning to predict by the methods of temporal differences. Mach. Learn. **3**(1), 9–44 (1988)
26. Volf, P.A., Willems, F.M.: A study of the context tree maximizing method. In: Proceedings of 16th Benelux Symposium on Information Theory, Nieuwerkerk Ijsel, Netherlands, pp. 3–9 (1995)
27. Watkins, C.J.C.H.: Learning from delayed rewards. Ph.D. thesis, King's College, Cambridge (1989)
28. Wilson, M.: Six views of embodied cognition. Psychon. Bull. Rev. **9**(4), 625–636 (2002)

Neural-Inspired Anomaly Detection

Stephen J. Verzi(✉), Craig M. Vineyard, and James B. Aimone

Sandia National Laboratories, Albuauerque, NM 87185, USA
{sjverzi,cmviney,jbaimon}@sandia.gov

Abstract. Anomaly detection is an important problem in various fields of complex systems research including image processing, data analysis, physical security and cybersecurity. In image processing, it is used for removing noise while preserving image quality, and in data analysis, physical security and cybersecurity, it is used to find interesting data points, objects or events in a vast sea of information. Anomaly detection will continue to be an important problem in domains intersecting with "Big Data". In this paper we provide a novel algorithm for anomaly detection that uses phase-coded spiking neurons as basic computational elements.

1 Introduction

Anomaly detection is an important research topic for data analysis [8,9], in general, and for image processing [1,6], in particular, as well as in its application to both physical security (for reduction of nuisance alarms) [10] and cybersecurity (in intrusion detection) [3]. The determination of anomalous data requires a contextual framework as well as a metric for comparison. In images, the x and y axes provide a spatial context, and pixel-to-pixel value differences are reasonable for comparison. With video (as in physical security) or streaming data (as in cybersecurity), the time dimension adds to the context, and other features as well as combinations of features are relevant for comparison.

In this research, we explore the suitability of spiking neural algorithms to address the problem of anomaly detection. In particular, we provide and analyze a new neural-inspired algorithm for performing noise-filtering on grayscale images that uses spiking neurons as fundamental elements for computation. This algorithm was specifically designed to operate using leaky integrate-and-fire (LIF) neurons [7] with phase-coded spiking activations [13]. Verzi et al. have shown that under certain conditions, e.g., symmetric distributions and data sets with unique sample median, a neural inspired algorithm can be optimal in the parallel random access machine (PRAM) computational complexity framework in comparison with other parallel algorithms [14]. Our paper expands upon previous work to show how to use multiple layers of phase-coded spiking neurons to implement anomaly detection in the process of noise-filtering of images.

2 Background

Noise filtering [5,12] is particularly well suited to the spiking algorithm presented in this paper, because the range of pixel values is typically bounded and small

A. J. Morales et al. (Eds.): ICCS 2018, SPCOM, pp. 202–209, 2018.
https://doi.org/10.1007/978-3-319-96661-8_21

(i.e., 0 through 255 for 8-bit grayscale) and the spatial context is easily represented using two-dimensional networks of neurons, resulting in an intrinsically parallel operation performance. Median-filtering has been used to filter noise from images for many years. One particular algorithm, adaptive center-weighted median-filtering (ACWMF) [2], has been successfully used to filter many different types of noise from images.

One benefit from using image processing and noise filtering to demonstrate an anomaly detection algorithm is that it is easy to visualize the difference in performance. In addition to showing how our algorithm differns from standard median-filtering and ACWMF, we will also provide metrics for comparing the performance of these algorithms. One useful metric is the number (or percentage) of pixels altered during processing, and another is the average change in pixel value, either across all pixels or only those pixels altered during processing. We will also provide computational complexity analysis and comparison of our algorithm with standard median-filtering.

3 Phase-Coded Spiking Neural Network

The spiking neural algorithm presented in this paper extends upon the one described by Verzi et al. [13]. Our multi-layer spiking adaptive median-filtering (AMF) network is shown in Figs. 1 and 2. The neural units in spiking AMF are LIF neurons which operate according to the following equation.

$$\begin{aligned} u_j(t) = {} & \left(u_j(t-1) - \lambda_j \left(u_j(t-1) - u^{\mathrm{eq}}\right)\right)\left(1 - z_j(t-1)\right) + u^{\mathrm{eq}} z_j(t-1) \\ & + \Sigma_{i=0}^{P} w_{ij} x_i(t) + \Sigma_{q=1}^{P} \rho_q \left(z_q(t-1)\right) \end{aligned} \tag{1}$$

In Eq. (1), u_j is the potential for neuron j; λ_j is the leakage rate; u^{eq} is the equilibrium (or resting) potential; z_j is 1 when neuron j fires (and 0 otherwise); x_i are real-valued inputs; z_q are phase-coded spike inputs with a phase-code value of ρ_q (from other phase-coded spiking neurons); and w_{ij} are input weights. In this paper, $u^{\mathrm{eq}} = 0$, and we use full leakage, $\lambda_j = 1$, to keep the algorithm simple. The phase-code for neuron j is computed as follows.

$$\rho_j \left(z_j(t)\right) = z_j(t) \left(t \bmod k\right) \tag{2}$$

In this paper, we use phase-coding to represent spiking delay directly within a phase-coding window of length k (phase steps). Thus, neuron j fires with a delay of $\rho_j/k = (t \bmod k)/k$. Phase-coding allows the spiking algorithm to process inputs and internal computations in ascending order, and including a delay of dk will delay the neuron from firing by $\lceil d/k \rceil$ phase windows plus $d \bmod k$ phase steps.

Spiking AMF is a simplification of ACWMF [2] where we do not include any center-weighting. Using phase-coding and our own notation, spiking AMF is described in the following equation set (for $m \in \{1, 2, 3, 4\}$).

$$o_{ij} = \begin{cases} \hat{\rho}_{ij}^{1}, & \exists m, \hat{\rho}_{ij}^{m} > \theta^{m} \\ x_{ij}, & \text{otherwise} \end{cases} \tag{3}$$

where

$$\hat{\rho}_{ij}^{m} = \underset{x \in \Omega_{ij}^{m}}{\text{median}}\{x\}$$
$$\Omega_{ij}^{m} = \{x_{lr} \mid i - m \leq l \leq i + m, j - m \leq r \leq j + m\}$$
$$\theta^{m} = s \cdot \underset{x \in \Omega_{ij}^{m}}{\text{median}}\left\{\hat{\rho}_{ij}^{m} - x\right\} + \delta^{m}$$
$$\delta^{m} = \frac{\left((2m+1)^2 - 1\right)}{2}$$

The values for θ^m are used to provide spiking thresholds for the spiking neurons r^m in Fig. 2, and in this part of the spiking network, the phase-coded but unprocessed pixel value u_{ij} is delayed by $5k$ or five phase-code windows to allow the appropriate time for anomaly detection in each of the neighborhoods. If there is no anomaly (i.e., none of the r^M neurons spike), then the original pixel value passes through unprocessed.

Figure 1 shows how input is converted from numbers to phase-coded spikes (input layer), how these phase-coded spikes are used to compute the median-value for each pixel within a particular neighborhood (median layer) and how the absolute difference from the median for each pixel is computed (absolute difference layer). Note that only one of the neurons, a_{ij} or a_{ji}, can fire as the other will have a negative overall input.

Figure 2 shows how the median of the absolute difference from the median (MAD) is computed for each pixel (median absolute difference from the median layer) and how the adaptive median-filter (AMF) value for each pixel is computed (adaptive median filter layer). Figure 1 and the left part of Fig. 2 show spiking AMF configured for a neighborhood of 3 × 3 (corresponding to Ω_{ij}^1 & r_{ij}^1) but our algorithm also includes neighborhoods of 5 × 5 (Ω_{ij}^2 & r_{ij}^2), 7 × 7 ($\Omega^3 ij$ & r_{ij}^3) and 9 × 9 (Ω_{ij}^4 & r_{ij}^4), as depicted on the right part of Fig. 2. Multiple iterations of spiking AMF are handled by routing the outputs (o_{ij}) back to the median-filter layer bypassing the input layer and continuing forward to the outputs again. In the next section we will describe the results of spiking AMF and compare it to standard median-filtering and ACWMF.

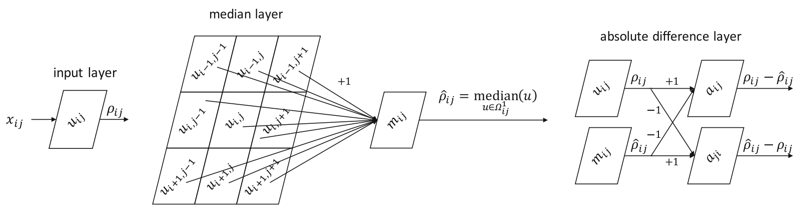


Fig. 1. Spiking architecture to translate input into phase, compute median and absolute differences.

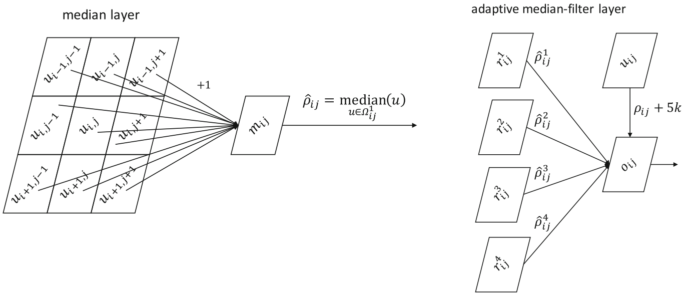


Fig. 2. Spiking architecture to compute median absolute difference from the median (MAD) for four increasing sizes of neighborhoods (3 × 3, 5 × 5, 7 × 7 and 9 × 9) and the spiking adaptive median-filter (AMF) output value (o_{ij}).

4 Results

To show benefits of spiking AMF, we compare it to standard median-filtering (MF) and ACWMF using an image common to the noise filtering research community [4,5,16,17]. In these results, each algorithm is iterated three times to produce a final, filtered image (see Figs. 3 and 4) from which the metrics of comparison are computed (see Table 1). For spiking AMF, we use $s = 3$, and for ACWMF, we use $s = 0.6$ and $\delta^m \in \{40, 25, 20, 5\}$ for random-valued and $\delta^m \in \{55, 40, 25, 15\}$ for fixed-value impulse noise as suggested by Chen and Wu [2]. Beyond these parameter settings, we did not attempt to optimize any of the algorithms. We also provide a simple computational complexity analysis and comparison for standard median-filtering and spiking AMF at the end of this section.

4.1 Examples 1 and 2 - "Pepper" and "Salt" Noise

In the first examples in Fig. 3, we show the standard MF algorithm implemented using our spiking framework [14] in comparison with our spiking AMF anomaly detection algorithm as well as ACWMF on an example image with 10% added "pepper" noise and the same image with 10% added "salt" noise. In MF, each pixel in the filtered image is computed from the original pixel and its surrounding neighbors (using a 3 × 3 neighborhood) using the median value from this neighborhood. This process tends to blur (almost all of) the pixels in the filtered image, as this algorithm does not specifically determine noisy from non-noisy pixels but rather smoothens the entire image. In spiking AMF as in ACWMF anomaly detection is used to determine which pixels are anomalous (or noise) versus those which should be kept unaltered.

4.2 Examples 3 and 4 - "Salt-and-Pepper" and "Random Impulse" Noise

In the second examples in Fig. 4, we compare all three algorithms on an example image with 10% added "salt-and-pepper" noise and the same image with 20% added "random impulse" noise. Table 1 shows pixel percent and value change comparison for the noise-filtering algorithms using the four noise cases presented in this paper. Spiking AMF results in a percentage of pixels changed, after three iterations, at the same magnitude as the noise introduced, which provides strong evidence that it is performing very precise anomaly detection. Both ACWMF and spiking AMF produce very low levels of average pixel value change.

Computational Analysis. Standard median-filtering has a computational (runtime) complexity of $O(N \log N)$ [17], and spiking AMF has a parallel computational (runtime) complexity of $O\left(kN/\sqrt{P}\right)$, where we assume there are N inputs and P neurons in our spiking neural framework. The analysis for spiking AMF borrows from Verzi et al. [13], where the addition of multiple layers adds a constant factor to the runtime and does not affect the asymptotic value presented here. The denominator ($\sqrt{P}$) is due to the fact that in the limit, N will be

Fig. 3. In the upper part, we compare all three algorithms with the addition of 10% "pepper" noise, and in the lower part there is 10% "salt" noise.

Table 1. Comparison of change in pixels for the noise-filtering algorithms (after three iterations) across four noise types.

Noise type	Percent of pixels changed				Average pixel value change: all (changed pixels only)			
	Noisy image	MF	ACWMF	Spiking AMF	Noisy image	MF	ACWMF	Spiking AMF
10% "pepper"	9.9	68.8	14.3	10.8	13.9 (140.8)	7.0 (10.2)	2.8 (19.7)	2.1 (19.7)
10% "salt"	10.0	68.9	14.5	11.0	11.8 (117.1)	7.1 (10.2)	2.9 (20.3)	2.0 (18.5)
10% "salt & pepper"	10.1	67.4	23.4	10.8	12.8 (127.5)	6.7 (10.0)	4.0 (17.2)	1.9 (17.9)
20% "random"	20.0	72.4	31.2	19.4	15.5 (77.7)	7.8 (10.8)	5.5 (17.5)	3.5 (18.1)

Fig. 4. In the upper part, we compare all three algorithms with the addition of 10% "salt-and-pepper" noise, and in the lower part there is 20% "random impulse" noise.

much larger than P, and we can only gain benefit from processing $\sqrt{P} \times \sqrt{P}$ size patches at a time. Reduced energy use is another important benefit of spiking AMF over non-spiking algorithms [13].

Future Work. Ongoing future work will continue to explore novel uses of our spiking neural framework as well as use of our anomaly detection algorithm in other application domains, such as physical security and cybersecurity. Neuromorphic anomaly detection in these domains is advantageous as it will provide reduced energy performance as well as intrinsic temporal context and computation, not just point disturbances. The temporal dimension is an important area of continuing research for data science [11,15]. An open problem in anomaly detection involves including temporal dimensions into the feature context and metric spaces.

5 Conclusions

In this work, we have demonstrated the utility of our neural-inspired algorithm, spiking AMF, to provide anomaly detection with specific demonstration in noise-filtering of images. This paper continues the current trend in research for developing energy efficient neuromorphic solutions. We have also provided evidence for the benefit of using neural-inspiration in designing computational solutions to interesting and important problems.

Acknowledgments. This paper describes objective technical results and analysis. Any subjective views or opinions that might be expressed in the paper do not necessarily represent the views of the U.S. Department of Energy or the United States Government. Sandia National Laboratories is a multimission laboratory managed and operated by National Technology and Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell International, Inc., for the U. S. Department of Energy's National Nuclear Security Administration under Contract DE-NA0003525. SAND No. 2018-5891 C.

References

1. Chang, C.I., Chiang, S.S.: Anomaly detection and classification for hyperspectral imagery. IEEE Trans. Geosci. Remote Sens. **40**(6), 1314–1325 (2002). https://doi.org/10.1109/TGRS.2002.800280
2. Chen, T., Wu, H.R.: Adaptive impulse detection using center-weighted median filters. IEEE Signal Process. Lett. **8**(1), 1–3 (2001). https://doi.org/10.1109/97.889633
3. Denning, D.E.: An intrusion-detection model. IEEE Trans. Softw. Eng. **SE–13**(2), 222–232 (1987). https://doi.org/10.1109/TSE.1987.232894
4. Dong, Y., Chan, R.H., Xu, S.: A detection statistic for random-valued impulse noise. IEEE Trans. Image Process. **16**(4), 1112–1120 (2007). https://doi.org/10.1109/TIP.2006.891348
5. Garnett, R., Huegerich, T., Chui, C., He, W.: A universal noise removal algorithm with an impulse detector. IEEE Trans. Image Process. **14**(11), 1747–1754 (2005). https://doi.org/10.1109/TIP.2005.857261
6. Honda, T., Nayar, S.K.: Finding "anomalies" in an arbitrary image. In: Proceedings of the Eighth IEEE International Conference on Computer Vision, (ICCV), pp. 516–523. IEEE (2001). https://doi.org/10.1109/ICCV.2001.937669
7. Jolivet, R., Lewis, T.J., Gerstner, W.: Generalized integrate-and-fire-models of neuronal activity approximate spike trains of a detailed model to a high degree of accuracy. J. Neurophysiol. **92**(2), 959–976 (2004). https://doi.org/10.1152/jn.00190.2004
8. Muthukrishnan, R., Poonkuzhali, G.: A comprehensive survey on outlier detection methods. Am. Eurasian J. Sci. Res. **12**(3), 161–171 (2017). https://doi.org/10.5829/idosi.aejsr.2017.161.171
9. Patcha, A., Park, J.M.: An overview of anomaly detection techniques: existing solutions and latest technological trends. Comput. Netw. **51**, 3448–3470 (2007). https://doi.org/10.1016/j.comnet.2007.02.001
10. Stubbs, J.J., Birch, G.C., Woo, B.L., Kouhestani, C.G.: Physical security assessment with convolutional neural network transfer learning. In: Proceedings of the 2017 International Carnahan Conference on Security Technology (ICCST). IEEE (2017). https://doi.org/10.1109/CCST.2017.8167800
11. Teng, H.S., Chen, K., Lu, S.C.Y.: Adaptive real-time anomaly detection using inductively generated sequential patterns. In: Proceedings of the IEEE Computer Society Symposium on Research in Security and Privacy, pp. 278–284. IEEE (1990). https://doi.org/10.1109/RISP.1990.63857
12. Tomasi, C., Manduchi, R.: Bilateral filtering for gray and color images. In: Proceedings of the Sixth International Conference on Computer Vision (IEEE Cat. No. 98CH36271), pp. 839–846. IEEE (1998). https://doi.org/10.1109/ICCV.1998.710815
13. Verzi, S.J., Rothganger, F., Parekh, O.D., Quach, T.T., Miner, N.E., James, C.D., Aimone, J.B.: Computing with spikes: the advantage of fine-grained timing (Accepted)
14. Verzi, S.J., Vineyard, C.M., Vugrin, E.D., Galiardi, M., James, C.D., Aimone, J.B.: Optimization-based computation with spiking neurons. In: Proceedings of the IEEE 2017 International Joint Conference on Neural Network (IJCNN), pp. 2015–2022. IEEE(2017). https://doi.org/10.1109/IJCNN.2017.7966098

15. Woo, B.L., Birch, G.C., Stubbs, J.J., Kouhestani, C.G.: Unmanned aerial system detection and assessment through temporal frequency analysis. In: Proceedings of the 2017 International Carnahan Conference on Security Technology (ICCST). IEEE (2017). https://doi.org/10.1109/CCST.2017.8167832
16. Yan, M.: Restoration of images corrupted by impulse noise and mixed Gaussian impulse noise using blind inpainting. SIAM J. Imaging Sci. **6**(3), 1227–1245 (2013). https://doi.org/10.1137/12087178X
17. Zhu, Y., Huang, C.: An improved median filtering algorithm for image noise reduction. Phys. Procedia **25**, 609–616 (2012). https://doi.org/10.1016/j.phpro.2012.03.133

Dynamics and Kinematics at Small Scales: From Micro and Nano Bubbles to Nanotubulation

Balakrishnan Ashok(✉)

Centre for Complex Systems and Soft Matter Physics,
International Institute of Information Technology Bangalore (IIIT-B),
26/C Electronics City, Hosur Road, Bangalore 560100, India
bashok1@gmail.com

Abstract. We discuss the behaviour of systems at the micro and nanoscales, looking at some interesting examples in particular. We first discuss our theoretical work on charged micro- and nano-bubbles undergoing radial oscillations in a liquid due to ultrasonic forcing. We obtain charge, frequency and pressure thresholds for the system. The electric charge affects the nonlinear oscillations of the bubble crucially, and limits the influence of the other control parameters such as the pressure amplitude and frequency of the driving ultrasound on the bubble dynamics. This has ramifications for medical diagnostics as well as industrial applications. We then report our theoretical work on the behaviour of nanotubes drawn out from micrometre-scale vesicles and exhibiting very interesting dynamics. Our theoretical model completely captures and reproduces all aspects of the force-extension curves reported in the experimental literature, completely explaining the dynamics of vesicular nanotubulation for the very first time.

Keywords: Bubble dynamics and cavitation · Nanotubulation
Nonlinear dynamics · Relaxation oscillations

1 Introduction

The behaviour of systems at small scales is an interesting field of study. The effects of spatial restriction can be unexpectedly manifested at the macroscopic level. Modelling and explaining such behaviour requires inclusion of a level of detail of the mechanisms involved at the micro and nanoscales. When the systems are driven or forced, the complexity involved increases and a rich dynamical behaviour may be displayed. We discuss here two examples of completely different systems at the micrometre and nanometre scales that exemplify such complex systems – the acoustically forced charged bubble undergoing radial oscillations in a liquid, and next, the dynamics of a vesicle that is pulled at a constant rate, causing the formation of nanotethers or nanotubes of uniform cross-section. Both systems are highly nonlinear, and display a wealth of complex behaviour.

A. J. Morales et al. (Eds.): ICCS 2018, SPCOM, pp. 210–219, 2018.
https://doi.org/10.1007/978-3-319-96661-8_22

For the charged bubble-oscillator, we find that limitations to the bubble's minimum radius also lead to limitations on the charge it can carry. We also obtain pressure thresholds for the bubble.

We discuss two cases here in both of which surface tension plays a major role in influencing the dynamics of the systems, even if in different ways. For the charged bubble, the presence of charge reduces the surface tension, and moves the system into a more periodic and ordered regime [3]. For the vesicle, the presence or absence of a lipid reservoir connected to the vesicle determines whether surface tension is kept constant or not; this appears in the model as effective spring constants, which in turn determines the dynamics of the system – whether the force-extension curves plateau or move gently upwards [4]. In both these small scale systems, the size and dimensions of the system play an important role and shows up in the observed dynamics of the system as a whole. For the vesicle, the stick-slip motion of individual phospholipid molecules on the vesicular wall translate into serrations in the force-extension $(f-x)$ curve for the entire vesicle itself. An increase in vesicular size also lowers the magnitude of the force in the $f-x$ curve. For the bubble, the magnitude of the maximum charge it can carry is again found to have a functional dependence on its equilibrium radius. Both these – the pulled vesicle and the charged bubble in an acoustic field – are examples of forced systems, where the amplitude of driving naturally plays an important role in influencing the dynamics, and where at small scales show up. Their study is intimately connected to the behaviour of living tissue under external forces.

2 Forced Oscillations of a Charged Bubble in a Fluid

2.1 A Brief Introduction

When an ultrasound pressure wave is directed at a gas bubble in a liquid, the bubble undergoes radial oscillations. The expansion and inward collapse of the bubble are violent events and the pressure thresholds where these happen are known as the Blake threshold and the upper transient pressure threshold, respectively. A gas bubble in water typically carries an electric charge on its surface due to the presence of disassociated ions in the liquid. To describe the behaviour of the bubble, the effects of charge Q can not be neglected as it reduces the effective value of the surface tension, and also affects the minimum radius the bubble can contract to, and its dynamics. The surface tension σ is effectively reduced to $\sigma - Q^2/(16\pi\epsilon_0 R_0^3)$ on introduction of charge. Making the assumption that the total charge on the bubble remains constant, the dynamics of the bubble are determined by modifying the Rayleigh-Plesset equation to include the effects of an electrostatic pressure term $Q^2/(8\pi\epsilon R^4)$, where $\epsilon = 85\epsilon_0$ and ϵ_0 is the electrical permittivity of vacuum. The equation is then given by [1–3]

$$
\begin{aligned}
&\left[\left(1-\frac{\dot{R}}{c}\right)R+\frac{4\eta}{c\rho}\right]\ddot{R}=\frac{1}{\rho}\left(P_0-P_v+\frac{2\sigma}{R_0}-\frac{Q^2}{8\pi\epsilon R_0^4}\right)\\
&\times\left(\frac{R_0}{R}\right)^{3\Gamma}\left(1+\frac{\dot{R}}{c}(1-3\Gamma)\right)-\frac{\dot{R}^2}{2}\left(3-\frac{\dot{R}}{c}\right)\\
&+\frac{Q^2}{8\pi\rho\epsilon R^4}\left(1-\frac{3\dot{R}}{c}\right)-\frac{2\sigma}{\rho R}-\frac{4\eta}{\rho}\left(\frac{\dot{R}}{R}\right)\\
&-\frac{1}{\rho}(P_0-P_v+P_s\sin(\omega t))\left(1+\frac{\dot{R}}{c}\right)-\frac{R}{\rho c}P_s\omega\cos(\omega t)
\end{aligned}
\tag{1}
$$

Here, R is the radius of the gas bubble with R_0 being its equilibrium radius, ω is the driving pressure frequency, P_s the amplitude of the driving sound pressure, P_0 the static pressure, P_v the vapor pressure, σ the surface tension, η the dynamical viscosity, ρ the density of the fluid, c the speed of sound in the fluid and Γ the polytropic exponent of the gas in the bubble.

2.2 Charge Thresholds

When a bubble is driven into radial oscillations by the applied pressure wave at driving angular frequency ω, we find that for the bubble's radial velocity to be some maximal value $\dot{R} = c_1$, a minimum magnitude of charge Q_{min} is required [1]. This Q_{min} is dependent on the driving frequency ω, and is given by

$$Q_{min}(\omega) \approx a(\omega - \omega_H)^{1/4}. \tag{2}$$

A plot of Q_{min} vs. ω therefore could be termed an isovelocity curve, demarcating a regime where $\dot{R} > c_1$ on one side, from $\dot{R} < c_1$ on the other. When the frequency equals a value ω_H, even an uncharged bubble of radius R would oscillate at that value of radial velocity c_1. ω_H depends on the pressure wave amplitude P_s through $\omega_H = b_1 + b_2 P_s$, where b_1 and b_2 vary with the radius of the bubble R_0. This dependence is found to be $\frac{d\omega_H}{dP_s} \sim R_0^{-0.9}$ [1].

We define a charge Q_h as the maximum charge a bubble can carry when the bubble radius reaches the van der Waals hard core radius h for the enclosed gas [1,2]. This maximal threshold charge Q_h also shows a dependence on ω. Indeed, the bubble dimensions affect its value as well, as could be expected. The actual dependence of Q_h on the equilibrium bubble radius is of the form [3]

$$Q_h \sim R_0^z. \tag{3}$$

We find that $z \approx 1.55$.

2.3 Pressure Thresholds

There are several pressure thresholds that affect the dynamics and growth of a gas-bubble in a fluid under acoustic forcing. The most well-known one is the

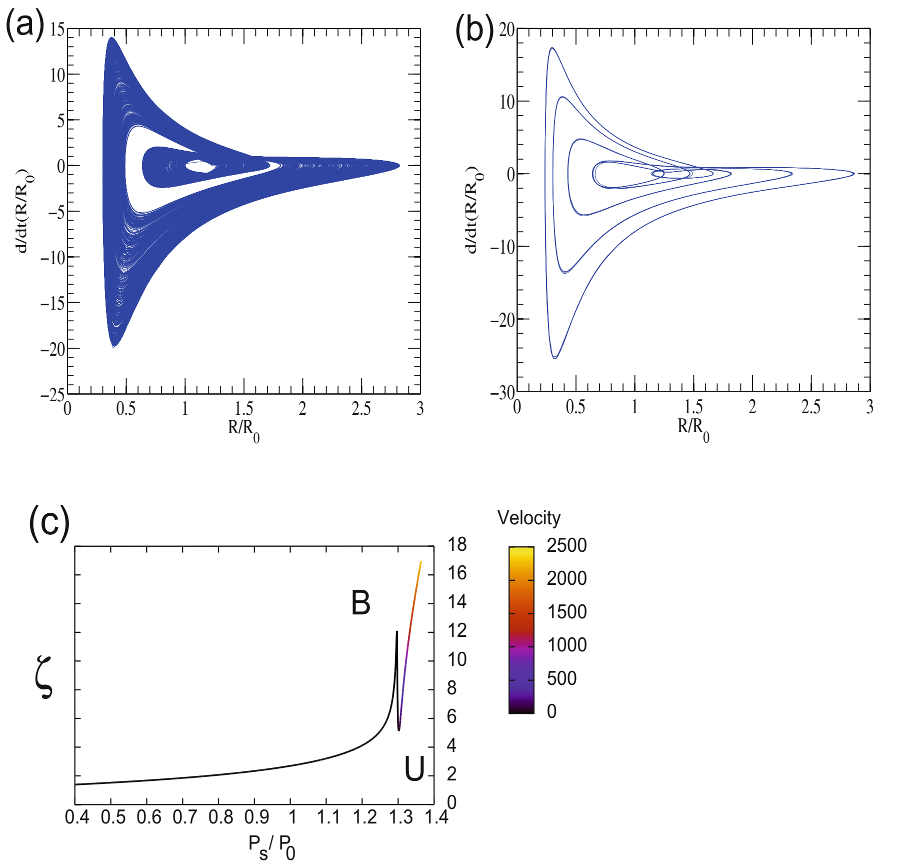


Fig. 1. (a) and (b) show typical $\dot{R}$ vs. R plots without and with charge, respectively. (c) shows an expansion-compression ratio plotted as a function of pressure. The spikes B and U correspond to the Blake threshold and the upper transient pressure threshold. The colour coded plot clearly shows that the bubble's radial velocity reduces as it expands to its maximum size at the Blake threshold, and increases to its highest values following collapse subsequent to the upper transient pressure threshold being passed.

Blake threshold, which is the lowest acoustic pressure exceeding which leads to explosive expansion of the bubble. The other is the upper transient pressure threshold, exceeding which causes violent bubble implosion. Electric charge changes the value of the thresholds. We find [1] the Blake threshold to be

$$p_{blake} = P_0 + \left(\frac{4\pi\epsilon\sigma^4}{Q^2}\right)^{1/3}\left(-6.95 + 4.146R_0\left(\frac{4\pi\epsilon\sigma}{Q^2}\right)^{1/3}\right) \tag{4}$$

An expansion-compression ratio ζ defined [1,2] through

$$\zeta \equiv (R_{max} - R_0)/(R_0 - R_{min}) \tag{5}$$

enables us to locate both the Blake threshold and the upper transient pressure threshold points on the same curve. Here R_{max} and R_{min} correspond to the maximum and minimum radius, respectively, achieved by the bubble during its forced oscillations.

Figure 1 (a) and (b) shows the effect of charge on the phase-plot for a bubble undergoing radial oscillations in a fluid – the difference is quite distinctive. The presence of charge clearly restricts the system to a narrower range of radius and velocities. Figure 1 (c) shows a typical expansion-compression ratio curve as a function of pressure, the Blake and upper transient pressure thresholds being marked on it. As can be seen, at pressures above the upper transient threshold, when violent bubble implosion takes place, higher radial velocities are achieved. The reader is referred to [1–3] for more details.

3 Dynamics of Vesicular Nanotubulation

3.1 Introductory Comments

The natural containers for chemicals in a living organism at the cellular level are vesicles, which consist of small phospholipid self-assembled structures, a few micrometres in dimension. The response of the vesicle to a constant pull velocity can be observed by fixing it at one end and pulling it at the other. When the fluid membrane of a vesicle is mechanically pulled at a constant velocity, it gets deformed and gets pulled out as a nanotube or tether with uniform cross-section. The dynamics and mechanics of this process has evoked a great deal of interest in the literature as the force-extension curves show several interesting features including steep rise and drop, as well as irregular fluctuations. While there had been many theoretical contributions on this subject, the dynamics had not been satisfactorily explained, and the literature had been singularly silent on the serrations. We showed through our theoretical model [4] that the fluctuations have a dynamical origin, in the stick-slip behaviour of the individual phospholipid molecules as they are pulled out by the external force. We successfully explained each and every feature seen in the experimental curves as consequences of various physical effects such as changes in vesicle curvature, surface tension effects, frictional forces, etc.

3.2 Features Observed in Force-Extension Curves

Experiments performed by Roopa et al. [5,6] and Cuvelier et al. [7,8] amongst others, on vesicles that have been pulled at constant velocity all show some characteristic, interesting features in their force -extension curves. These are typically an initial linear dependence of force on extension. This is then followed by a force drop (usually) then an irregular, serrated regime on a plateau which could either be horizontal or with a positive slope, the $f - x$ plot moving up gradually with time. Other features include an increase in the elastic threshold with increase in the pull velocity and a decrease in force values with increasing vesicle radius. All of these features are captured in the model explained below.

3.3 The Model of the Vesicular System

In experiments, the vesicle is either adhered to a substrate or to a pipette, and pulled at by a bead adhering to its surface. Our model simplifies the system, looking at the deformation of the vesicle over three separate stages as it is pulled, namely, its deformation from an idealized sphere to an ellipsoid, followed by the formation of a neck before tubule-formation, and lastly, tubulation and continued growth of the nanotubule with pulling (see Fig. 2). The neck geometry is idealized by modelling it as a cone of constant slope, and the nanotube as a cylinder of constant cross-section. Making the assumption that vesicular volume is conserved, from geometrical considerations we obtain [4] an expression for $R(t)$ the mean radius of the main vesicular body as a function of time as it is pulled out by a distance $x(t)$, R_0 being the initial radius:

$$R(t) \approx \left[{R_0}^3 - \frac{x(t)}{4} \left(r_0^2 + r_n(t)^2 + r_0 r_n(t) \right) \right]^{1/3}, \tag{6}$$

$r_n(t)$ is the radial cross-section at the top of the neck when in the neck-forming stage, r_0 is the radius of the base of the cone. When tubulation begins, we get

$$R(t) \approx \left[{R_0}^3 - \frac{3}{4} r_t^2 x - \frac{L_m}{4} \left(r_0^2 + r_t^2 + r_0 r_t \right) \right]^{1/3}, \tag{7}$$

with

$$r_n(t) = r_0 - \frac{x(t)}{L_m}(r_0 - r_t), \tag{8}$$

r_t being the radial cross-section of the nanotube. Taking into account the change in curvature as the sphere gets deformed into an ellipsoid, and also the subsequent increase in surface area, the force required to effect these deformations is

$$\begin{aligned} F_{ell} = \; & 16\pi\kappa \Big(R(t) - R_0 - x \Big) \left\{ \frac{\partial R}{\partial x} \times \left(\frac{R(t)}{(R_0 + x)(2R(t) - R_0 - x)^2} - \frac{1}{R_0^2} \right) \right. \\ & \left. + \frac{1}{R_0^2} - \frac{R(t)^2}{((R_0 + x)^2 (2R(t) - R_0 - x)^2)} \right\} + 4k_\sigma \pi \Delta A + k_\sigma \frac{x(t) r_0}{R(t)}, \end{aligned} \tag{9}$$

where k_σ is the surface-tension and ΔA is the change in area per unit length with respect to the sphere. During the second stage of neck formation, the additional point force that is required to pull out the vesicular shell by a distance L (upto a maximum of $L_m = \sqrt{2/3} r_0$, considering plastic flow for spheroidal shells) is obtained from elasticity theory to be

$$F_n = \frac{L^{1/2} E' h^{5/2}}{R(t)}. \tag{10}$$

Here, $E' = \frac{6}{5}(1 - \nu^2)^{-3/4} E$, and ν is the Poisson's ratio and E the Young's modulus; h is the thickness of the shell. Another effect of geometry is the significant curvature at the base of the neck as nanotubulation begins, due to expansion of the inner phospholipid monolayer and constriction of the outer monolayer. The corresponding force that needs to be overcome for a lipid molecule to be pulled past this bottle-neck, can be approximated by

$$F_{nc} \approx \alpha\pi\kappa\sqrt{4/15}/L_m, \tag{11}$$

$\alpha \geq 1$ being a numerical factor. Then for further deformation in the vesicle geometry leading to nanotube formation, accounting for the change in curvature energies, the force required is

$$F_{tube} = \pi\kappa\Big(\frac{1}{r_t} - \frac{4}{R_0}\frac{\big(1-\frac{R(t)}{R_0}\big)r_t^2}{R(t)^2}\Big). \tag{12}$$

The pulling rate influences the extent of internal flows as well as relaxation for the visco-elastic vesicular material. The rate-dependent response of the vesicular material can be modelled through a frictional force law, and a functional form is chosen that would support a velocity-weakening law in the neck-region, and that would generate a stick-slip process at the neck. For our work we chose a friction law of the form $F_{fr}[\dot{x}] = \frac{F_0\dot{x}/v_m}{(1+(\dot{x}/v_m)^2)}$. v_m is the value beyond which F_{fr} decreases with velocity $\dot{x}$. The forces acting at the neck and along the tube would also include a viscous resistance term $-\beta\dot{x}$, and purely elastic contributions from the stretching of the tubule and the vesicle. These forces can be written as

$$F_{e,f} = -\beta\dot{x} - \epsilon'[k'x + F_{fr}(\dot{x})], \tag{13}$$

$\epsilon' = 0$ for $x < L_m$ and $\epsilon' = 1$ for $x > L_m$. Here k', the effective spring constant of the tubule, can be weakening or constant with time, corresponding to either a flat serrated $f-x$ curve or a rising curve with non-zero, positive slope. The equations of motion for a bead of mass m that has adhered to the vesicle surface and which is being pulled at constant pull velocity v_a, can then be written as

$$\begin{aligned} m\ddot{x} &= f - F_n - F_{nc} + F_{e,f} - F_{ell} - F_{tube} \\ \dot{f} &= k_{eff}(v_a - \dot{x}). \end{aligned} \tag{14}$$

Here, k_{eff} is the effective spring constant of the vesicle. These equations are unstable for $v_a \geq v_m$.

Figure 3 shows an example of force-extension curves obtained from our model, and showing all the features seen in experimental plots. The source of each of the features of the curve have been marked on it.

More details may be found in [4].

4 Summary

We have briefly discussed some of our results on two small-scale systems – that of the oscillating charged bubble, and the vesicle yielding nanotubes on mechanical pulling. The bubble is typically of a few micrometres in dimension, as is the

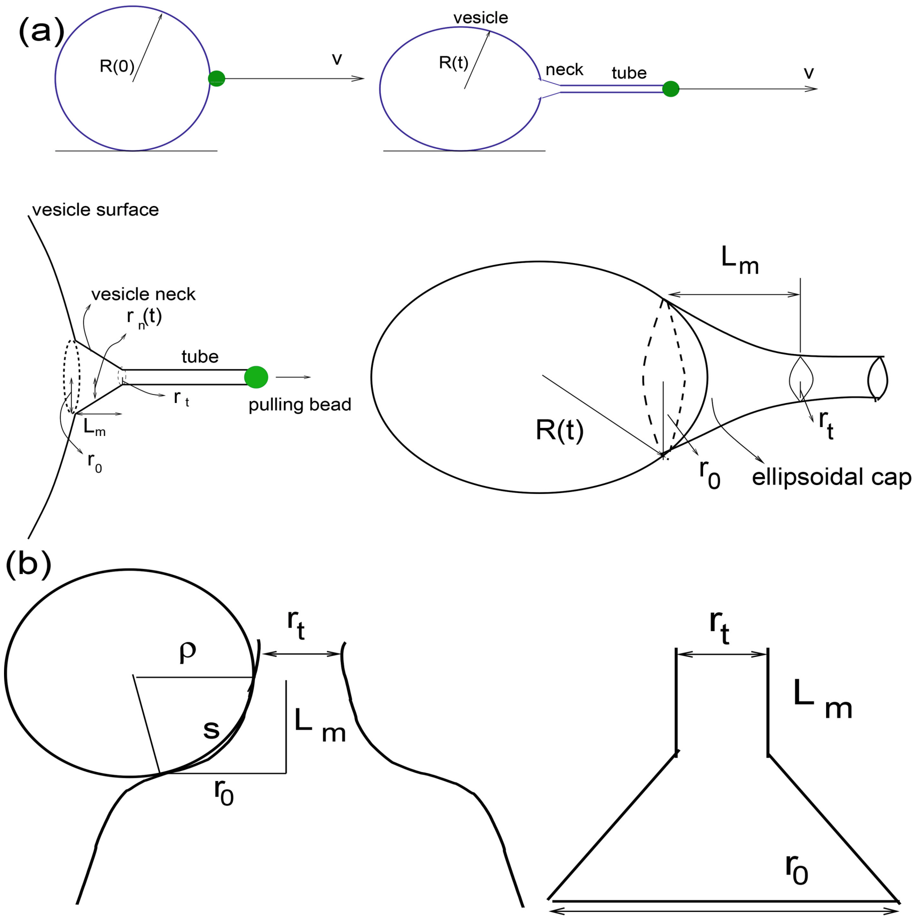


Fig. 2. (a) Schematic representation of the vesicular geometry. As each molecule is pulled out at the constricted neck, local curvature and friction has to be overcome. **(b)** For modelling, a cone-cylinder assembly is used to simplify the structure. The circle of radius ρ rolling over arc-length s corresponds to a lipid molecule being pulled across the enhanced local curved surface at the base of the neck (corresponding to a situation of increased curvature from κ to $\alpha\kappa$ in Eq. (11)).

vesicle. We see that in the case of the bubble, even the assumption of minute charge being present on its surface (typically of the order of 0.1 to no more than 10 pC) can not be neglected, and this has profound implications in its dynamics. Moreover, the fact that the bubble can not shrink in size indefinitely, and is bounded by the van der Waals hard core radius, conversely puts a limit on Q_h, the maximum magnitude of charge a bubble can carry physically and yields a power law dependence on R_0, the equilibrium bubble radius.

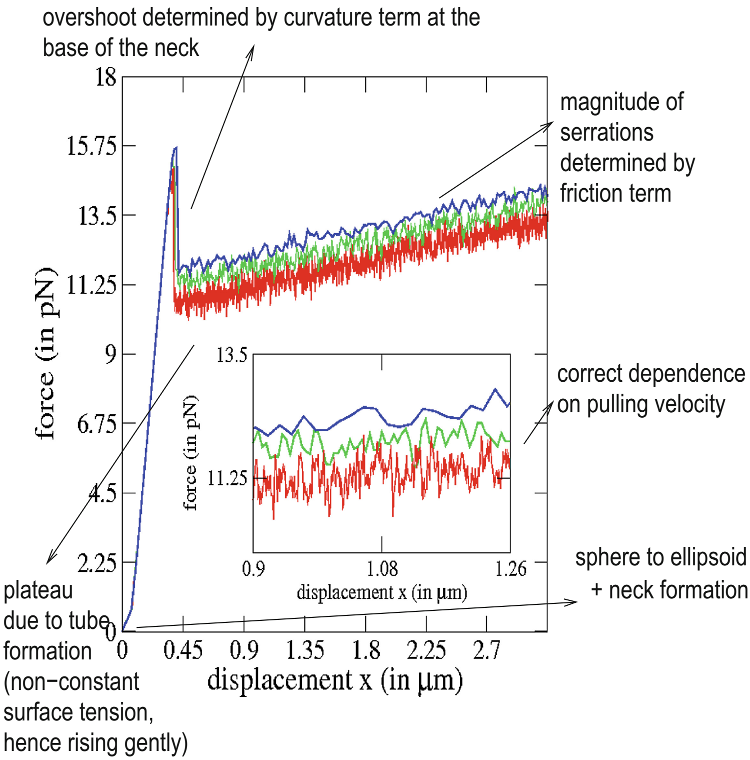


Fig. 3. Example of a force-extension curve obtained from our model (see [4]), showing the origins of the various features seen in nanotubulation curves. Note the change in slope at the base of the curve that can be attributed to the sphere-to-ellipsoid transformation and the subsequent neck formation. Plots shown are for three pull velocities: 2.17 µm/s (red), 21.72 µm/s (green), and 36.43 µm/s (blue). Increased pull velocities result in increased upper elastic limits. Smaller velocity curves lie below the larger ones. The mean period of oscillations is proportional to the pull velocity.

The nanotubulation of the vesicle shows how the motion of individual phospholipid molecules at the vesicular neck, and the frictional effect they have to overcome there, manifests itself macroscopically in the form of serrations seen in the force-extension $(f-x)$ curves. Moreover, we see how it is essential to account for local changes in geometry and curvature, in order to be able to reproduce the behaviour of the $f-x$ plots seen in experiments.

While the two systems are disparate ones, there is, nonetheless, a common thread connecting the two; vesicles can be used as micro-containers for localized, topical drug-delivery in a patient. The means of getting the vesicular walls to break or open to release their contents is by application of ultrasound to the localized part of the body. As living tissue is full of fluids and electrolytes, it becomes important therefore to understand the behaviour of charged bubbles on application of ultrasound, and these two studies – of bubble dynamics and vesicular nanotubulation, will, it is hoped be useful for this practical purpose as well.

Lastly a mention of our theoretical modelling of multi-walled carbon nanotubes would be relevant here. In a recent work [9], we have shown how a simple, theoretical, mechanical model reproduces results of interaction energies obtained from quantum-mechanical DFT calculations of double-walled carbon nanotubes. We also successfully predict the elastic constant for the system and the functional dependence of interaction energy on the dimensions of the nanotubes [9]. It is clear from our work that it is possible to successfully model the physics of structures at small scales using appropriate idealized models, and these can give us valuable insight into their behaviour.

References

1. Hongray, T., Ashok, B., Balakrishnan, J.: Effect of charge on the dynamics of an acoustically forced bubble. Nonlinearity **27**, 1157–1179 (2014)
2. Hongray, T., Ashok, B., Balakrishnan, J.: Oscillatory dynamics of a charged microbubble under ultrasound. Pramana J. Phys. **84**, 517–541 (2015)
3. Ashok, B., Hongray, T., Balakrishnan, J.: The charged bubble oscillator: dynamics and thresholds. Indian Acad. Sci. Conf. Ser. **1**, 109–115 (2017)
4. Ashok, B., Ananthakrishna, G.: Dynamics of intermittent force fluctuations in vesicular nanotubulation. J. Chem. Phys. **141** (2014). 174905
5. Roopa, T., Shivashankar, G.V.: Appl. Phys. Lett. **82**, 1631 (2003)
6. Roopa, T., Kumar, N., Bhattacharya, S., Shivashankar, G.V.: Dynamics of membrane nanotubulation and DNA self-assembly. Biophys. J. **87**, 974–979 (2004)
7. Cuvelier, D., Chiaruttini, N., Bassereau, P., Nassoy, P.: Pulling long tubes from firmly adhered vesicles. Europhys. Lett. **71**, 1015–1021 (2005)
8. Cuvelier, D., Derenyi, I., Bassereau, P., Nassoy, P.: Coalescence of membrane tethers: experiments, theory, and applications. Biophys. J. **88**, 2714–2726 (2005)
9. Mishra, B.K., Ashok, B.: Coaxial carbon nanotubes: from springs to ratchet wheels and nanobearings (2018, submitted, under review)

Rethinking Branch Banking Network

Oscar Granados(✉)

Department of Economics, Universidad Jorge Tadeo Lozano, Bogotá, Colombia
oscarm.granadose@utadeo.edu.co

Abstract. In recent years, banking services increased their digital services. However, they still require physical attention to customers and there may not be a definite extinction of the branches. Which is the best way to optimize the branch banking networks in megacities? This document proposes an alternative of branch banking network optimization, which uses genetic algorithms from information on the multi-layers structure of mobility, transportation, crime, cellular coverage, traffic and construction licenses. The results obtained define those locations where it may be more appropriate to establish a branch, as well as the need to merge or close other branches.

Keywords: Banking · City science · Complexity · Strategy

1 Introduction

The expansion of branch banking networks occurred during the second part of the Twentieth Century in areas of greater commercial mobility or political institutions, but not as much in those with highest population growth. The logic of branch networking was to develop, initially, in commercial areas with high mobility frequency or business centers. Later on, the evolution of the deregulation process changed the growth of branch networks [1–3]. Banking institutions have built extensive branch networks around cities to offer financial products and services, becoming part of the everyday life of people and the city's complexity [4, 5]. Therefore, banking has brought dynamical patterns to those locations where the population lives, works, agglomerates and interacts as a social structure [6, 7] because the presence of banks makes cities, and probably, creates the founding points on the spatial organization and articulation of production, services and markets [8].

How do banks define their branch network distribution? Scholars have different standpoints about the efficiency assessment of banks [8–10] and bank branches [11–17] with special attention to the DEA analyses. But, is the network a result of banking data? Can the banks shape human trajectories? [18] Have the banking networks been carrying out their expansion process under the urban dynamics logic or human mobility logic? To answer these questions, I proposed a complex system that integrates different elements that were never visualized to build a branch banking network and evaluates the evolution of these networks in megacities. My contribution through this work is to provide a means of measuring the best location for branches in a banking network based on city data criteria as a complement to the banking data. My focus is to explore a new framework of analysis to develop strategies for branch banking network

A. J. Morales et al. (Eds.): ICCS 2018, SPCOM, pp. 220–227, 2018.
https://doi.org/10.1007/978-3-319-96661-8_23

optimization, with an interaction among multilayer urban dynamics, to create more productive banks and efficient cities, i.e. urban reinvention in a complex future [19–21].

The motivations for this research are as follows. First, Bogota is one of the most important megacities in Latin America. Second, this document builds the efficiency of branches on the path to the digital banking transformation. Third, this analysis includes an assessment of the city's big data impact over the optimization model. With this motivation, this document has two closely related objectives. The first is to compute the interactions of city dynamical characteristics such as traffic, crime, pollution, transportation, mobile coverage and construction to identify the best business locations in the city. The second objective is to determinate the optimal branch location as result of these interactions and banking data. This document is organized as follows. Section 2 presents first, the dataset description to evaluate the characteristics of the city; and second, the dataset of the position of the 905 branches of ten unnamed Colombian banks. Section 3 provides the model, and Sect. 4 presents the results and the empirical findings. Finally, Sect. 5 provides concluding remarks and suggestions for further extensions of the work.

2 Data

The data such as transportation systems, urban traffic, construction licenses, crime, air quality, cellular coverage and cell phone tracing, provide patterns of urban dynamics and human mobility. These elements are potential sources of information for models of daily activities in a city, as is the banking service sub-system. Some works develop a similar proposal for other problems and objectives as an adaptive routing strategy [22–24] but this work seeks to optimize banking networks. Using datasets from the Digital Government of Colombia, the National Statistical System, the Local Government of Bogota, and data from telecommunications companies and the most relevant banking institutions, I show that my proposal allows for optimizing the branch banking networks in a megacity thanks to the synergetic effects between the urban dynamics and human mobility. I define first, branch banking network as a sub-system that adapts to the processes of the financial business and the interaction with the population, through their services (see Fig. 1).

In consequence, the branch networks of ten banks were identified in a city with 10 million inhabitants. Four of selected banks belong to the same local financial holding, but each one serves different market segments, the other three banks are subsidiaries of international banks, another bank is a partnership between a local group and an international bank, and the other two banks are mostly composed of local partners. Two local banks were founded in the 1870s, the others were founded during the second half of the 20th century. An international bank was established directly during the 1910s and the others were established through the acquisition of local banks in different years. The total branches of these banks in the city is 905, i.e., approximately 9.3 branches per 100,000 inhabitants or 14,4 per 100,000 banking customers.

Second, the datasets achieved to build six layers of multilayer methodology [25] correlate them all in the city of Bogotá. These layers are: The station layer, which represents the users of the transportation systems stations; the traffic layer, which

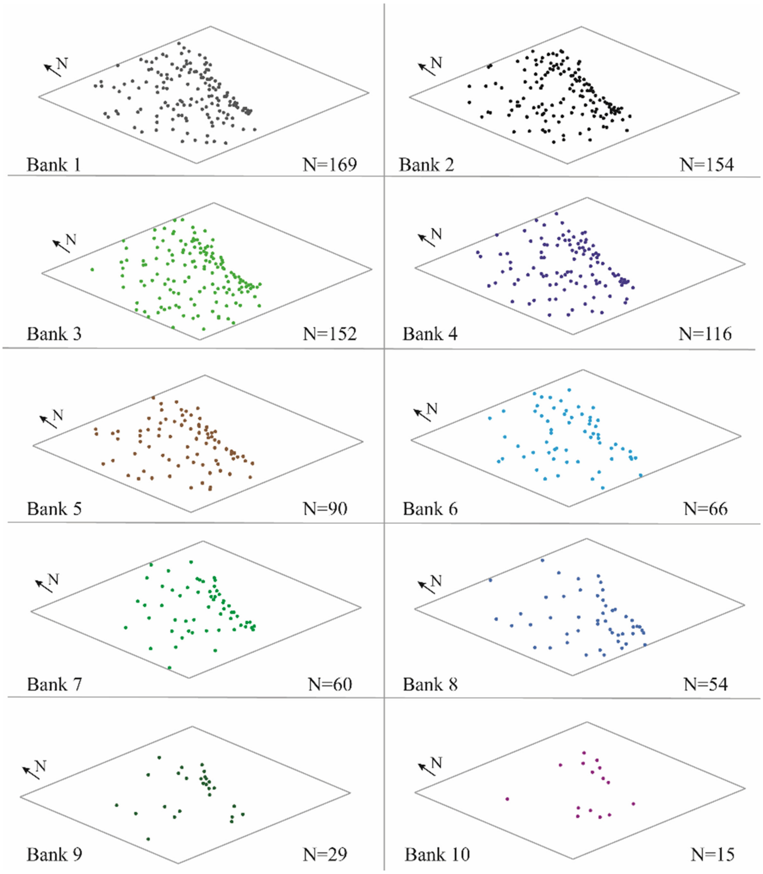


Fig. 1. Branch Banking Networks. The branches of ten unnamed Colombian banks are the points in the squared grid. This squared grid (B) represents the area of City of Bogota considered in this present research. The area is divided into a squared grid with 400 cells.

represents the average speed zones for public and private transportation; the construction layer, which is the result of the number of construction licenses by period; the crime layer, which represents the number of crimes in a period of twelve months; the air quality layer, which was generated from readings of twelve sensors scattered around the city, taken hourly over the course of twelve months; the cellular coverage layer, which was generated from the information of LTE coverage of the main mobile operators (see Fig. 2).

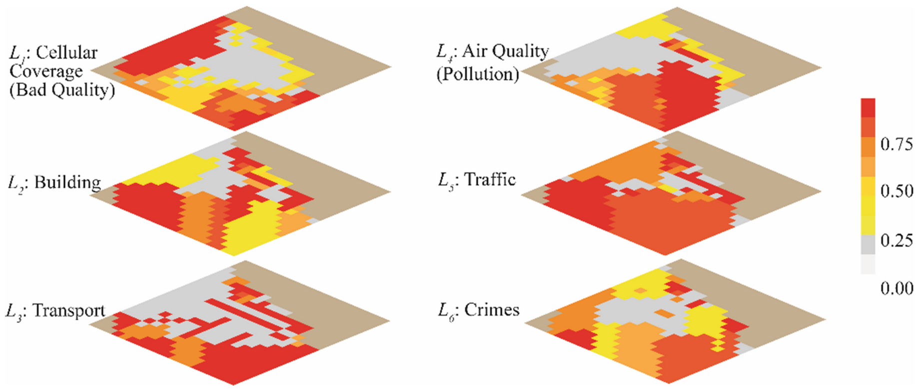


Fig. 2. City Layers. Each layer represents a restriction with different values for each one of 400 cells. Where the highest value is the red color and the lowest value is the gray color. The brown color is rural area.

3 Model

As for the methodological approach, the model of this research was composed of two parts: a city multilayer restriction analysis, and a genetic algorithm analysis. The first of these parts was designed to identify and map a few relevant features of city, and the second one allowed to identify an optimal branch banking network.

As a result, I consider Bogota an emerging market city, B and I discretize it into a grid of size $S \times S$. S is a squared grid. For the banking branches, I transport the addresses into geographical coordinates and I locate the branches into the grid without layers, as shown in Fig. 1. Layers L used the same grid B that represent the area of the City of Bogota excluding the city's rural areas. Some layers are static because this works as a restriction, especially for those that require the infrastructure construction and large temporal scales to change, such as cellular coverage, construction licenses or transportation. On the other side, some layers are dynamic, and correspond to restrictions with rapid changes such as traffic, air quality and crime. The restrictions were encoded by matrices $C^{\alpha}(t)$ and represented by layers $L_1, L_2, L_3, L_4, L_5, L_6$ (See Fig. 2) with different weighted combinations $w_{\alpha}(t)$ as shown in Eq. 1:

$$C(t) = \sum_{\alpha=1}^{n} w_{\alpha}(t) C^{\alpha}(t) \tag{1}$$

The weighted combinations allow to build the overall layer, but is important to establish that the static layers (L_1, L_2, L_3) are represented by $C_{\sigma}(s)$ and the dynamic layers (L_4, L_5, L_6) by $C_{\delta}(d)$. The weighting factor should be a function ranging between 0 and 1 accounting for the importance given to the different restrictions. The potential result layer represents an example of the weighted combination of static and dynamic restrictions as shown in Fig. 3a.

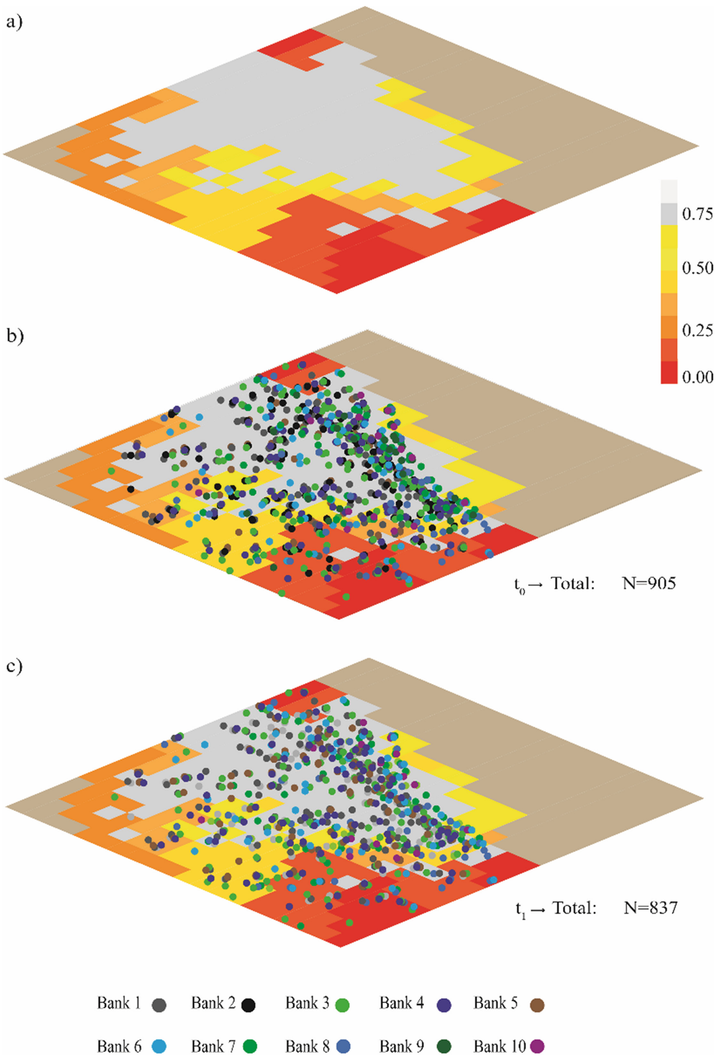


Fig. 3. Branch Banking Networks Optimization Results. (**a**) Potential result layer, as a result of weighted combination of L_1, L_2, L_3, L_4, L_5, L_6; where the high score location is the gray color and the low score location is the red color. (**b**) The square grid shows the result layer with 905 branches of the ten selected banks, without applying the algorithm. (**c**) On the bottom grid, the algorithm is applied considering the result layer, obtaining the branches that should be closed, merged or relocated. In total, the branch banking networks are adjusted by 68 branches optimizing the total network to 837 branches.

Additionally, the genetic algorithm is defined to optimize the branch banking network, i.e., to obtain the best number of branches by bank, the best location of new ones, the relocation of branches, or the best option to merge two or various branches. This intelligent optimization system allows to define a more precise solution for a complex optimization problem such as location [26–29] with dynamic and static layers. For this, the branch banking networks are coded with interactions that are generated from the layers, enabling a natural selection, where branch *b* is probabilistically eliminated, due to a lower adaptation to a complex scenario that emerges from the codified restrictions. The selected branches in a set of locations will be the least costly, determined by the restrictions in t_1 and their ability to survive and generate the next generation of branches in t_n. These branches will adapt to the guidelines that set forth the restrictions, and to the expanding of digital services.

Each banking branch in the city is ranked in increasing order of fitness, from 1 to *N*. The banks choose the expected value *Max* of the banking branch with rank *N*, with $Max > 0$. The expected value *Xv* of each banking branch *b* in the city at time *t* is given by:

$$Xv(b,t) = Min + (Max - Min)\frac{rank(b,t) - 1}{N - 1} \tag{2}$$

Where *Min* is the expected value of the banking branch with rank *1*. In each generation, the banking branches in the city are ranked and assigned expected values, according to Eq. 2.

4 Results

The results will be presented following the two sections shown in this research. I will first the potential result layer, developed for the integration of the weighted combination of static and dynamic restrictions; and then, I will discuss the optimization algorithm analysis carried out to evaluate the relations between the city data and the bank data, as well as different options that emerged to optimize the branch banking networks.

In the first phase, with the incorporation of different city data, and defining the weight of each layer, it was possible to assign a score to each of the 400 cells. This score, which may vary according to the weight given to each layer depending on the research interest, was the result of assigning equal weights to the six layers. However, with small adjustments, not exceeding a 5% range in each layer, the results did not have a significant deviation. The result layer shows a few low scoring locations that allow banks to determine how important it is to be or not, in this location, or to look for alternatives such as the construction of banking service clusters with other banks (ex.: shopping centers). Additionally, locations with high scores allow banks to direct their strategy by market segment, making the network more efficient.

In the second phase, I made interpolations between the available current data to identify patterns and mechanisms of current branch networks, and I performed simulations of branch banking networks to gain insights on which restrictions influence the location to define a new branch, a branch relocation, or a merger of two or various branches. These simulations allow for researching future plausible evolutions of the branch banking networks. As result, I consider that each bank can define the best strategy to create its branch network optimization as of the city's transformation, digital service evolution, and human mobility (see Fig. 3) because mobility and city design are becoming increasingly interdependent and interrelated [5]. Second, the application of the algorithm optimized the network by 68 closures (Bank 1: 12, Bank 2: 17, Bank 3: 9, Bank 4: 8, Bank 5: 4, Bank 6: 5, Bank 7: 7, Bank 8: 3; Bank 9: 3, Bank 10: 0) and 10 relocations. The closures are the result of branches very close to each other or a particularly low-score location, and relocations are a result of branches with better cell score within the same area. These results open perspectives for the future strategy of branch banking networks in megacities because, in terms of complexity, it would be interesting to understand the effect of having branch banking networks in cities with populations over 10 million [30].

5 Discussion and Conclusions

I have presented a strategy to define the best option for branch banking networks. This document used a multilayer structure and a genetic algorithm to define the optimization intelligence and show how to determine locations for the unnamed ten banks, and their exact number of branches, considering the city's evolution and the digital service transformation. The conclusions of this research can be summarized as follows. First, the document provides a better understanding of the characteristics of branch banking networks for ten unnamed banks in a city with 10 million inhabitants. Second, some banks depend on their business proposal upon low-scoring locations. Third, banks must develop corporate banking and private banking centers for network optimization. Fourth, banks must drive the largest number of branches for retail banking and small and medium-sized company services. Fifth, banks can have consolidated clusters between different banks in low score locations. Sixth, the application of this strategy allows to banks define the optimal location for their branch network with the least impact on their clients, in line with the digital transformation growth. Nevertheless, certain cautions are in order here. This empirical analysis looks at city data and banking data only for the year 2017. The banking business is determined by short term economic conditions and may vary from year to year, yet, decisions on the number of branches need to be based on multiple year data. Finally, the innovation progress may change the branch personnel with new ATMs or other AI technologies, an option to reduce costs and create an efficient banking network.

References

1. Carlson, M., Mitchener, K.: Branch banking as a device for discipline. J. Polit. Econ. **117**(2), 165–210 (2009)
2. Berger, A.N., Demsetz, R.S., Strahan, P.E.: The consolidation of the financial services industry: causes, consequences, and implications for the future. J. Bank. Financ. **23**, 135–194 (1999)
3. Dick, A.: Nationwide branch banking and its impact on market structure, quality and bank performance. J. Bus. **79**, 567–592 (2006)
4. Miller, J.H.: A Crude Look at the Whole: The Science of Complex Systems in Business, Life, and Society. Basic Books, New York (2016)
5. Bettencourt, L., West, G.: A unified theory of urban living. Nature **467**(21), 912–913 (2010)
6. Barabási, A.-L.: Linked: How Everything Is Connected to Everything Else and What It Means for Business, Science, and Everyday Life. Basic Books, Boston (2002)

7. Watts, D.: Six Degrees. The Science of a Connected Age. Norton & Company, New York (2003)
8. Friedmann, J.: The World City Hypothesis. Dev. Change **17**(1), 69–83 (1986)
9. Berger, A., Humphrey, D.: Efficiency of financial institutions: international survey and directions for future research. Eur. J. Oper. Res. **98**(2), 175–212 (1997)
10. Berger, A., Hanweck, G., Humphrey, D.: Competitive viability in banking: scale, scope, and product mix economies. J. Monet. Econ. **20**(3), 501–520 (1987)
11. Berger, A., Humphrey, D.: Measurement and efficiency issues in commercial banking. In: Griliches, Z. (ed.) Output Measurement in the Service Sectors. University of Chicago Press, Chicago (1992)
12. Lozano, S.: Slacks-based inefficiency approach for general networks with bad outputs: an application to the banking sector. Omega **60**, 73–84 (2016)
13. Wanke, P., Maredza, A., Gupta, R.: Merger and acquisitions in South African banking: a network DEA model. Res. Int. Bus. Financ. **41**, 362–376 (2017)
14. Jayo, M., Diniz, E., Zambaldi, F., Christopoulos, T.: Groups of services delivered by Brazilian branchless banking and respective network integration models. Electron. Commer. Res. Appl. **11**(5), 504–517 (2012)
15. Hirtle, B.: The impact of network size on bank branch performance. J. Bank. Financ. **31**, 3782–3805 (2007)
16. Sherman, H.D., Gold, F.: Bank branch operating efficiency: evaluation with data envelopment analysis. J. Bank. Financ. **9**(2), 297–315 (1985)
17. Schaffnit, C., Rosen, D., Paradi, J.: Best practice analysis of bank branches: an application of DEA in a large Canadian bank. Eur. J. Oper. Res. **98**(2), 269–289 (1997)
18. Paradi, J.C., Zhu, H.: A survey on bank branch efficiency and performance research with data envelopment analysis. Omega **41**(1), 61–79 (2013)
19. Gónzalez, M., Hidalgo, C., Barabási, A.-L.: Understanding individual human mobility patterns. Nature **453**, 779–782 (2008)
20. Bettencourt, L., West, G.: Bigger cities do more with less. Sci. Am. **305**(2), 52–53 (2011)
21. Glaeser, E.: Triumph of the City: How Our Greatest Invention Makes Us Richer, Smarter, Greener, Healthier, and Happier. Penguin Books, New York (2012)
22. Johnson, N.: Simply Complexity. One World Publications, Oxford (2007)
23. Widhalm, P., Yang, Y., Ulm, M., Athavale, S., Gonzalez, M.C.: Discovering urban activity patterns in cell phone data. Transportation **42**(4), 597–623 (2015)
24. De Domenico, M., Lima, A., González, M., Arenas, A.: Personalized routing for multitudes in smart cities. EPJ Data Sci. **4**(1), 2–11 (2015)
25. Florez, M., Jiang, S., Li, R., Mojica, C., Rios, R., González, M. (2016). humnetlab.mit.edu/wordpress/wp-content/uploads/2016/03/bogotatrb_2017.pdf. Accessed Jan 2018
26. Kivela, M., Arenas, A., Barthelemy, M., Gleeson, J., Moreno, Y., Porter, M.: Multilayer networks. J. Complex Netw. **2**, 203–271 (2014)
27. Davis, L.: Handbook of Genetic Algorithms. International Thompson Computer Press, Boston (1996)
28. Cortinhal, M.J., Captivo, M.E.: Genetic algorithms for the single source capacitated location problem. In: Metaheuristics: Computer Decision-Making. Applied Optimization, vol. 86, pp. 187–216. Springer, Boston (2003)
29. Kung, L.-C., Liao, W.-H.: An approximation algorithm for a competitive facility location problem with network effects. Eur. J. Oper. Res. **267**(1), 176–186 (2018)
30. Jaramillo, J., Bhadury, J., Batta, R.: On the use of genetic algorithms to solve location problems. Comput. Oper. Res. **29**(6), 761–779 (2002)
31. Cohen, J.: Human population: the next half-century. Science **302**(5648), 1172–1175 (2003)

Empirical Mode Decomposition in Defence Data Analysis

Peter Dobias[1(✉)] and James A. Wanliss[2]

[1] DRDC CORA, 60 Moody Drive, Ottawa, ON K1A0K2, Canada
peter.dobias@drdc-rddc.gc.ca
[2] Department of Physics and Computer Science, Presbyterian College, 503 South Broad Street, Clinton, SC 29325, USA
jawanliss@presby.edu

Abstract. Repetitive, but not strictly periodic, trends in the temporal data can present a challenge to the analysis of short-term patterns. Military examples of such time-series include violence data or vessel detections. Empirical mode decomposition (EMD) has been used across a variety of different fields such as biology and plasma physics to deal with non-stationarities in the data. This methodology enables separation of different modes intrinsic to the data and it does not require a priori assumptions about time dependence of various data sub-components, such as periodicity of variations. We show the application of this methodology to two distinct types of data. The Afghanistan violence data provided an example of a sparse, limited dataset. With EMD we were able to identify a multi-year cycle, without the skewed trend in the vicinity of turning points. In contrast, ship detection data for the Canadian West coast provide an example of a large data set. Unfortunately, the analysis of the summary detection data led to the presence of noise limiting our ability to identify specific patterns in the data. The analysis could be improved by geographically dividing the data into a number of small areas and conducting separate analysis for each area. Despite this, the EMD demonstrated its usefulness and applicability, enhancing the analysis of these two datasets compared to more conventional approaches.

Keywords: Empirical mode decomposition · Non-stationarities
Time series analysis

1 Introduction

Complex data, composed of multiple modes, are inherently difficult to analyze. There is a variety of methods looking at the identification of patterns in imagery or radio signal using various means of machine learning (e.g., [1] and references therein). This particular area is beyond the scope of this paper. Another example is time series capturing dynamics of various phenomena in biological, physical, or social systems. Such data often exhibit repetitive, but not strictly periodic, trends in the temporal domain; the presence of near-periodic cycles can present a challenge to the analysis of short-term patterns. Understanding of these patterns is important in order to be able to identify correlations with various underlying drivers and to facilitate at least short term

A. J. Morales et al. (Eds.): ICCS 2018, SPCOM, pp. 228–236, 2018.
https://doi.org/10.1007/978-3-319-96661-8_24

predictive analysis. Among the military examples of such time-series are security incident data, such as roadside bombs and ambushes [2], or vessel detections [3]. For instance, the analysis of weekly and monthly variations in Afghanistan violence data was hampered by the presence of annual fighting cycle (ranging from 11 to 13 months) [4]. In some cases such long-term trends can be removed in order to enable the analysis of short-term dynamics; one such methodology employed on Afghanistan data was multiplicative decomposition [2, 5, 6]. However, this methodology assumes exact periodicity, and consequently did not work close to the turning points [4]. Another example of annual cycles with slight variations from the 12-month period is the seasonal aggregations of fishing vessels in particular areas [5, 7]. While the migrating fish aggregate in particular areas around the same time, this migratory cycle can be influenced by a variety of factors such as predator prey dynamics, climatic changes, weather, sea currents, etc. [8].

Empirical mode decomposition (EMD) [9] has been used across a variety of different fields such as biology and plasma physics [10] to deal with non-stationarities in the data. This methodology enables separation of different modes intrinsic to the data and it does not require a priori assumptions about time dependence of various data subcomponents, such as periodicity of variations. Thus it provides a superior approach to conventional methods of analyzing violence data [4]. In addition, it provides a venue for the in-depth analysis of multi-scale dynamical properties of the data, such as self-correlation or intermittency [11, 12].

In this paper we revisit the earlier analysis of the violence data in Afghanistan (for completeness), and then we focus on the analysis of the temporal trends in the shipping data off the West Coast of Canada. Understanding of long, and short-term trends in the data can assist in the analysis of the pattern of life (POL) of shipping; this in turn can help with both detection of anomalies for security and defence agencies, long-term search and rescue (SAR) planning [13, 14].

The paper is organized as follows. In the next section we describe the EMD, and its applicability to the defence and security data. Then we summarize, for the sake of completeness, previously published analysis of the security incident data in Afghanistan. Finally, we present the analysis of the shipping data off the coast of Canada.

2 Empirical Mode Decomposition

Data describing dynamics of most real systems, whether natural or man-made, are often multi-modal, and exhibit nonlinear, often repetitive while not strictly periodic, behaviour. This presents a significant challenge for the analysis, and necessitates employing different analytical methods. Such methods should be capable of representing the inherent complex multiscale nature of dynamical systems [10–12].

Traditional approaches in analysing military data rely on *a priori* assumptions about the nature of the data, such as annual cycles. In reality, an *a priori* defined function should not be used to build such a basis, as it can lead to the circular argument about the nature of the data and the processes driving the data structure [6].

A limited number of adaptive methods are available for signal analysis, among them being empirical mode decomposition (EMD) [9, 15], which serves as a complement to Fourier and wavelet transforms. The EMD method employs transformation basis defined *a posteriori*, from the decomposition method derived from the data; it is intuitive and highly adaptive [6, 10]. The essential idea of the EMD is that any time series can be written as the superposition of a small number of monocomponent signals called intrinsic mode functions (IMFs), each characterized by a well-defined frequency [10]. Unlike for Fourier analysis, the functional basis for the EMD is intrinsic to the analyzed time series; since the modes are not necessarily periodic, it works for repetitive time series without well-defined periods.

The EMD is obtained through a sifting algorithm [6, 10]: Let $\{t_j\}$ be the local maxima of a signal $X(t)$. The cubic spline $E_U(t)$ connecting the points $\{(t_j, X(t_j))\}$ is referred to as the upper envelope of X. The lower envelope $E_L(t)$ is similarly obtained from the local minima $\{s_j\}$ of $X(t)$. Then we define the operator $\mathbf{S}$ by

$$S(X) = X - \frac{1}{2}(E_U + E_L). \tag{1}$$

In the so-called sifting algorithm, the first intrinsic mode function is given by

$$I_1 = \lim_{n\to\infty} S^n(X). \tag{2}$$

Subsequent intrinsic mode functions in the EMD are obtained recursively by

$$I_k = \lim_{n\to\infty} S^n(X - I_1 - \ldots - I_{k-1}). \tag{3}$$

The process stops when $Y = X - I_1 - I_2 - \ldots - I_m$ has at most one local maximum or local minimum. This function $Y(t)$ denotes the local trend in $X(t)$.

3 Security Incidents in Afghanistan 2006–2012

Dobias and Wanliss [4] analyzed security incidents in Afghanistan between 2006 and mid-2012 (Fig. 1, left); in their study they compare use of the multiplicative seasonal decomposition [2, 3] and the EMD to remove the near-annual trends from the data in order to determine short term variations. In order to facilitate discussion, we are including a brief summary of the findings here.

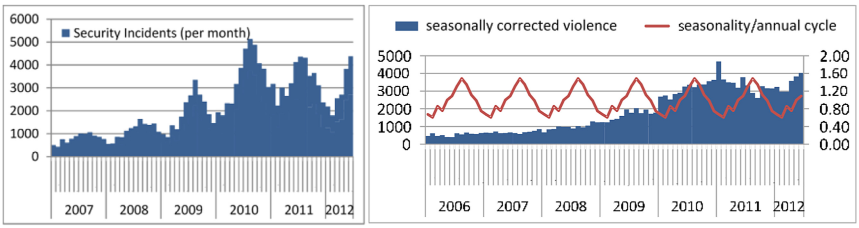


Fig. 1. Monthly security incident counts in Afghanistan 2006–2012 (left), and the seasonally decomposed data (right) (reproduced from [6]).

The multiplicative seasonal decomposition assumed a priori a 12-month cycle in the data; this was imposed as a multiplicative factor on the data (Fig. 1, right, red line). The seasonally corrected data (Fig. 1, right, blue) could then be analyzed for both short and long-term trends. For instance, in reference [2] the authors used residual analysis to extract variety of trends at several scales. However, because the annual cycle was imposed on the time series, the decomposed data showed some behaviour around the turning points that was the result of the cycle generally differing from exactly 12 month due to external factors such as weather (shorter or longer winter, rain), or human-imposed (deployment rotations, elections, or various legitimate or illicit harvest).

In contrast, the EMD (Fig. 2) did not assume a particular periodicity of the data. Instead, it yielded a set of base functions. These were then analyzed using Fourier transformation [4] to determine possible periodicity. This analysis revealed a set of repetitive behaviours; the annual one was actually varying between 10 and 13 months with the 12-month period being somewhat dominant. The latter finding justified post hoc the use of the seasonal decomposition apart from the turning points. Furthermore, there were also multi-year cycles identified in the data, as well as a number of multimodal, short term cycles [4]. Of a particular interest was the fact that while the short-term (days) spectrum was consistent with white noise, the medium to short-term spectrum (weeks to months) exhibited 1/*f* noise consistent with the power-law distribution [4], reinforcing earlier findings [16–18].

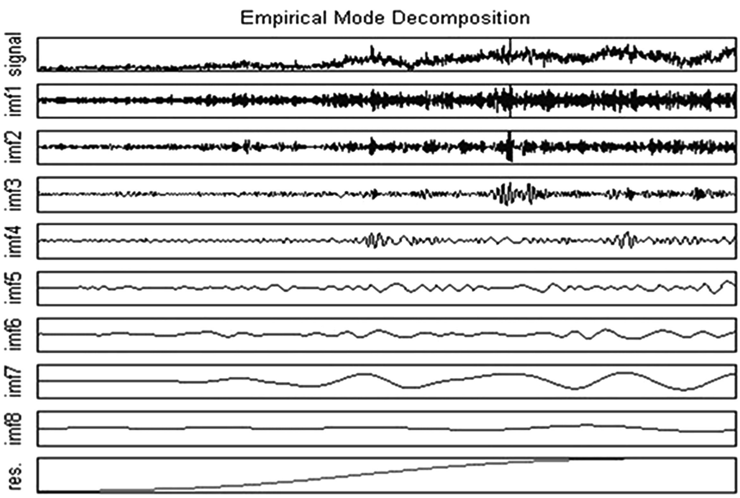


Fig. 2. EMD of the security incidents in Afghanistan 2006–2012 (reproduced from [6]).

Overall, the paper demonstrates that the EMD can significantly enhance the analysis of violence data and provide additional insights hidden from other methodologies [4].

4 Pattern of Life Analysis of Marine Shipping off the West Coast of Canada

Understanding the pattern of life (POL) for shipping traffic in the coastal areas is very important for law enforcement, national defence, and search and rescue (SAR) planners. Especially if the POL exhibits particular recurring trends, these could be used to efficiently manage the resources dedicated to surveillance, enforcement, and SAR throughout the year. For example, knowing that during a particular part of the year there is an increased shipping density in certain areas would enable planners to allocate needed resources to cope with any contingency situations. Understanding expected behaviour also facilitates detection of potential anomalous behaviour [13, 14].

In this paper we address the POL off the coast of western Canada. The primary source of the data is the unclassified Global Position Warehouse (GPW) database; the primary sensor feeding into this database is the Automatic Identification System (AIS), mandated under the UN Convention on the Law of the Sea (UNCLOS) for any vessels over 300 GT, and any passenger-carrying vessels. The sources of AIS data are coastal stations (that capture AIS near coast) as well as satellite-based AIS that have global reach. In addition, many smaller vessels use AIS as well to facilitate safety and navigation. The AIS data is supplemented by other sensors such as coastal radars, aerial surveillance assets, and satellite imagery. Due to diversity of sensors, these will likely have a spatio-temporal structure of their own; this structure might be desirable to compare with the distribution of the detections. Once again, we have to reiterate that we are not interested in identifying a specific pattern in any of the particular sensors; rather, the purpose of our analysis is to identify overall patterns in temporal distribution of the detections across all sensors.

While some of the trends can be estimated qualitatively from the general knowledge of the area, its climatic conditions and weather, and marine life migration patterns, these are typically rather coarse and do not maximize the benefits that can stem from detailed understanding of the POL patterns and their drivers. Hence we attempt to employ EMD to decompose the shipping data and identify spatio-temporal trends in these data. The work presented here is preliminary, and represents a proof-of-concept work rather than a comprehensive spatio-temporal analysis.

For the initial analysis we used all of the detections within the area spanning the coast of West Canada, and reaching approximately 1000 km off shore. Overall there were approximately 6 million detections within that timeframe (Fig. 3).

The EMD (Fig. 4) exhibits rather noisy behaviour in 2017–18. The long-term component (residual) suggests that following a steady slow drop in the recorded detection between 2014 and 2016 there was a fairly steep increase in the recorded data between 2017 and 2018; this may have contributed to the observed noise in the data. The shorter-term modes show occasional spikes. Since there was no apparent periodicity observable in the data, we conducted spectral analysis of the nine lowest modes.

The spectral analysis (Fig. 5) confirmed the noisy nature of the data across all scales. There appears to be very limited repetitive periodicity; most modes exhibit wide power spectrum that suggest some repetitive behaviour ranging mostly between 1 and 4 months, and then spanning multiple years.

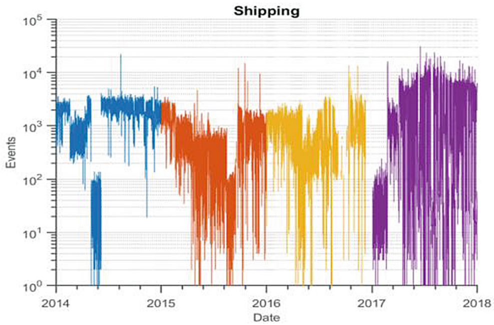


Fig. 3. Ship detections off shore of western Canada between 2014 and 2018 binned by hour.

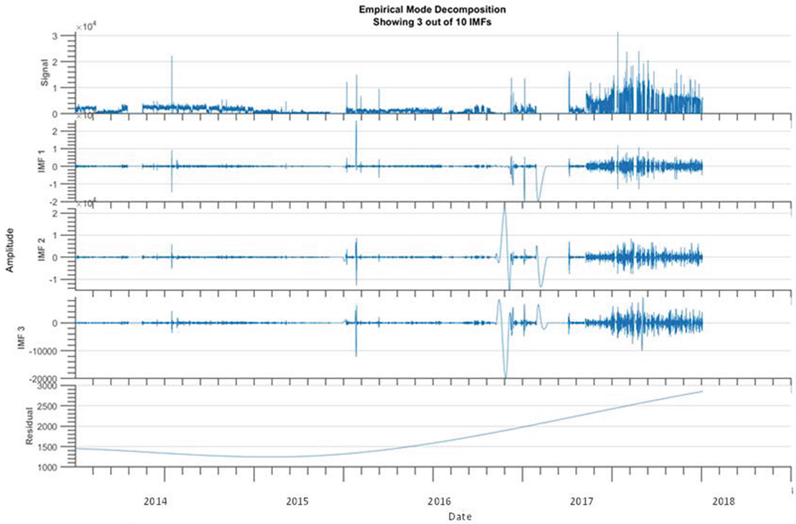


Fig. 4. Ship detections off shore of western Canada between 2014 and 2018 binned by hour.

However, the behaviour overall is dominated by noise. Interestingly, the noise does not appear entirely random, and may exhibit some 1/f dependence typical for fractal systems. This would be consistent with earlier analysis by Liu et al. [13]. However, more detailed, and more geographically focus analysis would be required. Conducting the analysis on a refined spatial grid might also lead to the reduction of the noise caused by temporal shifts in increased traffic density between different areas.

In summary, the current analysis demonstrated a feasibility of using the EMD to analyze temporal patterns in vessel traffic. However, in order to achieve better understanding of these patterns, more spatially refined analysis would be necessary. Using the summary data for a large geographical area led to the presence of significant noise that seems to suppress any finer patterns. Hence, while there is some

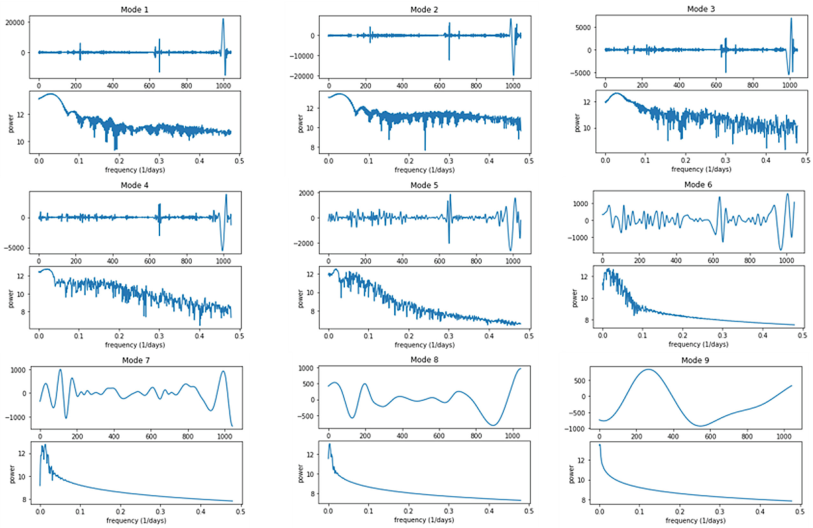


Fig. 5. Spectral analysis for the nine lowest modes; the upper of the picture pair shows the IMF while the lower shows frequency dependence of the power.

repetitiveness in the data, the clarity of observations at the level required for detailed, especially predictive, POL analysis was impossible to achieve.

5 Summary and Conclusions

While the use of EMD for the analysis of violence data from Afghanistan provided some valuable insights into the conflict dynamics and temporal patterns in violence across the country for the analyzed period, the analysis of maritime shipping data was not as successful. In order to fully benefit from the EMD for the latter, the data will need to be divided into refined geographical grid to constraint analysis to limited areas. This will enable us to avoid introducing cross pollution of the trends by different timings of traffic density changes in different areas (e.g., slightly different times for fishing seasons in different zones). In addition, it will likely be beneficial to analyze track densities rather than detection densities to avoid problems with different sensor coverage of different areas. However, the impact of the latter will likely be somewhat reduced if the spatial grid is fine enough.

Overall, we have demonstrated that the EMD enables separation of different modes intrinsic to the data and it does not require *a priori* assumptions about time dependence of various data sub-components, such as periodicity of variations. Thus it provides a superior approach to conventional methods of analyzing violence data. In addition, it provides a venue for the in-depth analysis of multi-scale dynamic properties of the data, such as self-correlation or intermittency. We show the application of this methodology to two distinct types of data. The Afghanistan violence data between 2009 and 2013 provided an example of a relatively sparse, limited dataset. With EMD we were able to identify a multi-year cycle, without the skewed trend in the vicinity of turning points. In addition, the EMD isolated shorter time scales with distinct statistical behaviour thus enriching the opportunities for analysis of drivers at different scales. In contrast, ship detection data for the Canadian West coast provide an example of a very large data set. In order to analyze patterns of life in certain areas, a discrete density function was developed, and then the individual points were analyzed separately. A combination of analyses of cumulative and binned density functions provide insights into seasonal trends in the ship detections (and arguably shipping). The understanding of the various scales on which the shipping density varies will enhance the ability to predict the future likelihood of finding particular vessel types in certain areas, thus improving the efficiency and effectiveness of employing different sensors. Overall, the EMD demonstrated its usefulness and applicability, enhancing the analysis of these two datasets compared to more conventional approaches.

References

1. Wiskott, L., Sejnowski, T.J.: Slow feature analysis: unsupervised learning of invariances. Neural Comput. **14**(4), 715–770 (2002)
2. Gons, E., Schroden, J., McAlinden, R., Gaul, M., VanPoppel, B.: Challenges of measuring progress in Afghanistan using violence trends: the effects of aggregation, military operations, seasonality, weather, and other causal factors. Defence Secur. Anal. **28**(2), 100–113 (2012)
3. Dobias, P., Horn, S., Liu, M.-J., Eisler, C., Sprague, K.: Use of fractal-based approaches in the assessment of the Canadian recognized maritime picture. In: Proceedings from 32nd International Symposium on Military Operational Research, London, UK (2015)
4. Dobias, P., Wanliss, J.A.: Empirical mode decomposition applied to Afghanistan violence data: comparison with multiplicative seasonal decomposition. J. Battlefield Technol. **17**(2), 17–22 (2014)
5. Dobias, P., Eles, P., Schroden, J., Wanliss, J.: Quantitative analysis in support of military intelligence: predictive analysis and risk assessment in current security environment. In: Proceedings from the 28th ISMOR, Hampshire, UK (2011)
6. Wanliss, J.A., Dobias, P.: Dealing with non-stationarities in violence data using empirical mode decomposition. J. Battlefield Technol. **16**(2), 2–5 (2013)
7. Radtke, H.D., Davis, S.W.: Description of the US West Coast commercial fishing fleet and sea food processors, Report for Pacific States Marine Fisheries Commission (2000). www.psmfc.org/efin/docs/fleetreport.pdf. Accessed 10 April 2018
8. Binder, T.R., Cooke, S.J., Hinch, S.G.: The biology of fish migration. In: Farrell, A.P. (ed.) Encyclopedia of Fish Physiology: From Genome to Environment, vol. 3, pp. 1921–1927. Academic Press, San Diego (2011)
9. Huang, N.E., Shen, Z., Long, S.R., Wu, M.C., Shih, S.H., Zheng, Q., Tung, C.C., Liu, H.H.: The empirical mode decomposition and the hilbert spectrum for nonlinear and non-stationary time series analysis. Proc. R. Soc. A-Math. Phys. **454**, 903–993 (1998). https://doi.org/10.1098/rspa.1998.0193
10. Yu, Z.-G., Anh, V., Wang, Y., Mao, D., Wanliss, J.: Modeling and simulation of the horizontal component of the geomagnetic field by fractional stochastic differential equations in conjunction with empirical mode decomposition. J. Geophys. Res. **115**(A10219) (2010). https://doi.org/10.1029/2009ja015206

11. Anh, V., Yu, Z.-G., Wanliss, J.A.: Analysis of global geomagnetic variability. Nonlin. Process. Geophys. **14**(6), 701–708 (2007). https://doi.org/10.5194/npg-14-701-2007
12. Wanliss, J.A., Weygand, J.M.: Power law burst lifetime distribution of the SYM-H index. Geophys. Res. Lett. **34**(L04107) (2007). https://doi.org/10.1029/2006gl028235
13. Liu, M.-J., Dobias, P., Eisler, C.: Fractal patterns in coastal detections on approaches to Canada. J. Appl. Oper. Res. **7**(2), 80–95 (2015)
14. Horn, S., Eisler, C., Dobias, P., Collins, J.: Data requirements for anomaly detection. In: Proceedings from Maritime Knowledge Discovery and Anomaly Detection, Italy (2016)
15. Shi, F., Chen, Q.J., Li, N.N.: Hilbert Huang transform for predicting proteins subcellular location. J. Biomed. Sci. Eng. **1**, 59–63 (2008)
16. Dobias, P.: Self-organized criticality in asymmetric warfare. Fractals **17**(1), 91–97 (2009)
17. Sprague, K., Dobias, P., Bryce, R.: Critical behaviour in patterns of violence in Afghanistan. J. Battlefield Technol. **14**(1), 17–24 (2011)
18. Dobias, P., Wanliss, J.A.: Fractal properties of conflict in Afghanistan revisited. J. Battlefield Technol. **15**(3), 31–36 (2012)

Scenario Set as a Representative of Possible Futures

Peter Dobias(✉) and Cheryl Eisler

DRDC CORA, 60 Moody Drive, Ottawa, ON K1A0K2, Canada
{peter.dobias, cheryl.eisler}@drdc-rddc.gc.ca

Abstract. Because of the inherent disconnect between the length of the procurement cycle and the dynamic time scales of the security environment, future military capability planning presents a unique challenge. The security environment can change very rapidly, and the extent of the changes, and thus the future environment, is potentially unpredictable. In contrast, the procurement cycles responsible for the renewal of the capabilities are typically on the order of decades. The traditional approach to capability-based acquisition consists in the development of a scenario set representing possible futures, and then plan capabilities to address the challenges presented by the scenario set. This approach is based on an implicit assumption that the chosen scenario set is sufficiently representative of all possible futures. Intuitively, if the scenarios are selected somewhat randomly (i.e., there is no significant selection bias that would eliminate certain futures), then a solution that works for the set of planning scenarios, whether in optimization or capability scheduling sense, is more likely to work for any feasible future. In this paper we examine this assumption in a more systematic way and attempt to express likelihood that a solution to a limited set of vignettes is a solution for all possible futures, at least for some special cases.

Keywords: Planning scenarios · Modeling and simulation · Scenario selection

1 Introduction

Because of the inherent disconnect between the length of the procurement cycle and the dynamic time scales of the security environment, future military capability planning presents a unique challenge [1]. The security environment can change very rapidly, and the extent of the changes, and thus the future environment, is potentially unpredictable [2]. In contrast, the procurement cycles responsible for the renewal of the capabilities are typically on the order of decades. The traditional approach to capability acquisition consists in the development of a scenario set that is supposed to represent possible futures, and then plan capabilities to address the challenges of the scenario set [3, 4]. This approach is based on an implicit assumption that the scenario set is sufficiently representative of all possible futures, i.e., that a solution to this scenario set will be a solution for any of the feasible futures. A similar approach, with a similar set of assumptions, appears in almost any type of future planning, where the optimized solution to unknown future is sought. Intuitively, if the scenarios are selected somewhat

A. J. Morales et al. (Eds.): ICCS 2018, SPCOM, pp. 237–244, 2018.
https://doi.org/10.1007/978-3-319-96661-8_25

randomly (i.e., there is no significant selection bias that would eliminate certain class of futures), then a solution that works for more of the planning scenarios, whether in optimization or capability scheduling sense, is more likely to work for any feasible future [2]. However, even for moderately complex systems, the number of possible scenarios can be very large (or even infinite) [5].

There are multiple factors that drive the complexity of the problem at hand. Furthermore, individual scenarios can be influenced at a variety of scales. For example, at the geopolitical/strategic level, a set of possible force planning scenarios can be identified [6]; since the future security environment cannot be predicted with certainty there is an inherent uncertainty in the scenario selection. For each of the feasible (plausible) future scenarios a variety of operational responses can be envisioned [4, 7]. At the operational level, scenarios can be further parametrized with different scope of the problem (e.g., size of involved units) and various weapon and sensor characteristics in order to produce vignettes with tactical level responses. Lastly, tactical level planning can be fine-tuned with detailed, physics-based models that provide highly realistic simulation capabilities for weapons and sensor performance.

In this paper, we attempt to look at the scenario selection and the resulting comprehensiveness of the representation of the future in a systematic way without resorting to massive scenario generation [8]. The strategic level (development of possible futures and their likelihood) requires in-depth intelligence and strategic analysis and is not addressed in this paper. However, the operational and tactical levels are addressed. We look at the issue of selecting a set of operational vignettes, and determining critical confidence bounds on parametric uncertainty based on historical source data. Then we look at the tactical-level modeling, and especially at the issue of parameterization of such tactical scenarios (e.g., modeling a range of sensor or weapon capabilities). We look at the potential impact of a selecting a too limited sample on the potential conclusion that would be driven from an analysis.

The paper is organized as follows. First, an example of the operational level confidence assessment using historical data is reviewed. In the real-world analysis this step would set scene for the lower (tactical to low operational) level modeling and analysis. An example of the latter is then addressed through an artificial vignette using parametric modeling in an agent-based model. Typically, a number of vignettes would be required to represent each strategic scenario. However, the one example here suffices for the demonstration purposes. Finally, the results and their implications are discussed in the context of potential future work for capability development.

2 Parameterization of Force Planning Scenarios

Government policy, departmental strategy and priorities, and an assessment of the security and defence environment are top-down inputs that feed into the development of a set of a strategic force planning scenarios [7]. These scenarios are expected to represent the entire range of possible missions that a force structure will be expected to accomplish. Individual scenarios are quantified with a specific set of capability requirements. The force structure is then evaluated to determine if it can meet the capability requirements for all scenarios [9]. A key step is to characterize the types of missions along parameters of interest and ensure that the appropriate range of activities is covered off in the solution set. Scenarios are often categorized by scope, scale, role, threat level, likelihood, consequence of failure, urgency, expected duration, and capabilities required. Ordinal or categorical ranking may provide sufficient fidelity for some parameterization.

In order to provide suitable evidence that scenarios are realistic, justifiable, and sufficiently representative as a solution for any of the feasible futures, historical evidence can be used to align past events with future planning scenarios in a bottom-up approach. In some cases, it may be necessary to subdivide planning scenarios into operational-level vignettes that either provide more detail than the strategic-level scenarios would ordinarily give or may be capability-specific. For example, a scenario may focus on international emergency assistance in general, while vignettes for further analysis may be developed for the assistance in particular locations.

While historical data is not strictly a predictor of future events and likely to miss-estimate the likelihood of extreme events [10], it does provide a valuable starting point for estimation of probability distributions for likelihood. Taking the example for flood relief, incidents of severe flooding can be tracked over time. Assuming that the time between event occurrences happens according to a Poisson distribution, then the count of events (n) over a given interval (i) provides the mean frequency estimate (λ). Given a confidence level (α), the upper (λ_{ul}) and lower (λ_{ll}) central confidence intervals on the mean can be estimated using the Neyman procedure (for a small number of events), where $(X^2)^{-1}$ is the inverse chi square distribution [11].

$$\lambda_{ll} = \frac{(X^2)^{-1}\left(\frac{1+\alpha}{2}; 2n\right)}{2i} \lambda_{ul} = \frac{(X^2)^{-1}\left(\frac{1-\alpha}{2}; 2(n+1)\right)}{2i} \tag{1}$$

Thus, a probabilistic estimate for the likelihood of the vignette can be estimated with no prior knowledge of the sample population. In the cases where scenarios are projecting forward to events that are dissimilar from all prior events, then subject matter expert opinion is often sought to provide some justification for the likelihood of a potential vignette. Vignette durations can also be constructed using probabilistic methods. There are a variety of methods that could be applied, including random sampling with replacement, simple two- or three-point estimation, or distribution fitting using the maximum likelihood estimator procedure [12]. Once the maximum likelihood vignettes are estimated, they can be translated into tactical vignettes that can be in turn used to evaluate system performance for both human and technology aspects, introducing additional dimensions to the scenario sampling problem.

3 Ship Protection Against Asymmetric Threats: Parametric Modeling

To address the impact of sampling on the outcome for a tactical scenario, we have used a simple naval scenario of a vessel's force protection against small boat swarms [13]. The modeled scenario considers a single frigate, moving at slow speed through a narrow navigation channel. The remainder of the area was assumed to be shallows of insufficient depth for the frigate to traverse, while imposing no limitations on small boats (civilian or threat). Such limitation is common in many coastal areas. We have varied several parameters (specifically, the number of attackers, type, range and effectiveness of weapons). Then we selected several subsets of obtained results, and compared them to the overall results obtained across all modeled options.

The scenario implemented an agent-based model (ABM) called Map-Aware Non-linear Automata (MANA) [14]. As an ABM, MANA represents a rather abstracted version of reality. It is intended to analyze emergent behaviour across a large number of repetitions rather than to provide an absolute assessment of systems' effectiveness. MANA entities (agents) can be assigned a variety of characteristics (speed, sensors, up to six weapon systems, armour protection, communication capabilities, etc.). These can differ between different agent states; for example, some sensors can be activated only if an agent is fired upon, or if it received information about a possible threat from another agent. MANA agents move within terrain features represented by an RGB (red-green-blue) colour map of a bitmap file. Different colours represent various levels of protection, ease of movement, and concealment. The distances are represented in "points" (pts) which are, in effect, pixels of the terrain bitmap [15].

The scenario was implemented as follows. A Blue (friendly) vessel moved along a straight line across a 200 $\times$ 200 pts square area of water. No avoidance manoeuvres were considered, as those would risk grounding the vessel [13]. Between 1 and 10 Red (enemy) fast attack craft approached the Blue ship, using the civilian traffic (modeled as 25 randomly slow-moving neutral vessels) as a cover. Once in the vicinity of the target, the Red vessels would accelerate to get within range to trigger on-board explosives. Blue would warn any vessel crossing a predetermined distance threshold, with approximately 80% probability that the message was received. The "refuel" capability of MANA was used to model the communications; this was done earlier for crowd modeling [16]. Upon being refueled, the agents would change their state. In the new state the civilian boats would turn away from Blue, while the Red boats would accelerate toward Blue. Blue had lethal (a gun firing in bursts of 6 rounds) and non-lethal capabilities available. While the range of the lethal capability was further than the warning point, Blue would not employ it against targets outside of the warning range unless they were prior identified as hostile (and of appropriate threat level). The non-lethal system was a generic system that could represent either a warning device or a radio-frequency engine stopper, as its primary role was to convey a "do not approach" message to the target vessel. The ranges and effective area of the non-lethal system were parameterized (range from 10–70 pts in increments of 10 pts, and the effective radius from 5–50 pts in increment of 5 pts). The parametric analysis was completed for 3, 5, 7 and 10 Red attackers. Blue would use non-lethal capability first; if the boat continued approaching, then lethal force would be employed. It was also assumed that two explosions were required to kill the Blue frigate; hence, a single Red boat could, at best, damage the target.

First, we examined the variability of the number of Red attackers while keeping weapons systems parameters fixed (Blue range equal to 75 pts for the lethal system, with approximately 30% kill probability for a burst of six rounds, and a non-lethal system with a range of 30 pts and the effects radius 25 pts). Table 1 shows the results.

Table 1. Dependence of the Blue hit/kill on the number of attackers.

Number of Red	Blue killed		Blue hit		Red killed		Civilians killed	
	Mean	Sigma	Mean	Sigma	Mean	Sigma	Mean	Sigma
1	0.00	0.000	0.21	0.041	1.0	0.000	0.00	0.000
2	0.08	0.027	0.44	0.050	2.0	0.000	0.00	0.000
3	0.21	0.041	0.47	0.050	3.0	0.022	0.00	0.000
4	0.34	0.048	0.44	0.050	3.9	0.035	0.00	0.000
5	0.48	0.050	0.42	0.050	4.6	0.066	0.02	0.014
6	0.64	0.048	0.31	0.046	5.2	0.098	0.01	0.010
7	0.77	0.042	0.20	0.040	6.0	0.126	0.01	0.010
8	0.85	0.036	0.14	0.035	6.4	0.150	0.00	0.000
9	0.93	0.026	0.06	0.024	6.6	0.185	0.04	0.020
10	0.95	0.022	0.05	0.022	7.0	0.174	0.00	0.000

The probability of Blue getting killed increases along a sigmoid curve $y = 1.199 - (0.005 + 1.199)/\left(1 + \frac{x^{2.548}}{83.150}\right)$ with R^2 = 0.998. The middle section (between 2 and 7 attackers) can be approximated by a linear function. While other measures were available, since the purpose of the study was to look at the dependence of scenario on the parameterization rather than an actual operational advice, we determined that a single measure was sufficient to examine here.

Hence, if either of the extremes (1 attacker, or 9–10 attackers) are removed, the information about the diminishing return is lost. We have selected several subsets of the full set of various attacker numbers ({2,3,4,5,6,7}, {1,2,3,4,5}, {1,3,5,7,9}, {2,4,6,8}, and {6,7,8} attackers). The slope for the trend in Blue killed varied from 0.09 for the upper part of the estimate {6,7,8} attackers to 0.14 for the lower part of the curve. Selecting every other option provided very similar results to the overall trend for the entire considered range. Similar observations can be made for the Blue hit numbers as well, except the saturation point is reached sooner (for approximately five attackers), and the variability in the slope is even greater, from 0.01 to 0.16.

Varying the range of the non-lethal system yielded a series of heat maps (the results for the number of Blue kills) shown in Fig. 1. Clearly, the number of Red attackers was the main driver, the range and area coverage had very little impact on the results. However, the relative randomness of the results implies that the findings could be somewhat dependent on the selection of the range/radius combination. For 3, 5, 7, and 10 attackers the distribution of the results is normally distributed (using Shapiro-Wilk normality test [17]); the mean and standard deviation (σ) values are in Table 2. Selecting several sub-samples with limited weapons effects radius (25 and 50 pts. vs. 10, 20 and 30 pts., as well as ranges 20, 30 and 40 pts. versus 50, 60 and 70 pts) revealed that, in this case, the results are completely independent of the sampling. In all the cases, the mean was consistent with the results across all the parameters. So in this instance the scenario sampling made no difference, and any sample would provide the same answer for each of the selected numbers of attackers.

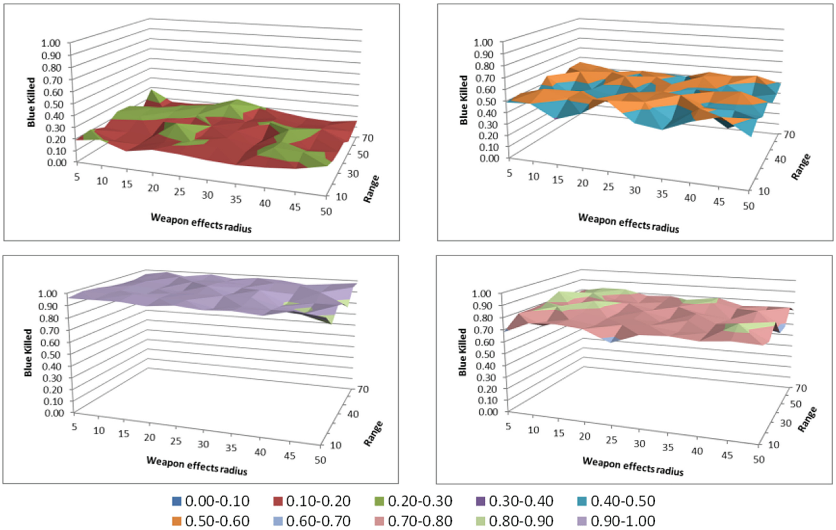


Fig. 1. Dependence of the number of Blue killed on the range and area coverage of the non-lethal system for (clockwise, starting top left) 3, 5, 7, and 10 attackers.

Table 2. Mean number of the Blue killed across all ranges and effects radii for different numbers of attackers.

Number of attackers	All options		All ranges	All ranges	All radii	All radii
	Mean	+/− σ	Radius (pts) (5,25)	Radius (pts) (10,20,30)	Range (pts) (20,30,40)	Range (pts) (50,60,70)
3	0.19	0.04	0.19 +/− 0.03	0.19 +/− 0.04	0.19 +/− 0.04	0.19 +/− 0.04
5	0.49	0.05	0.51 +/− 0.04	0.50 +/− 0.05	0.50 +/− 0.05	0.49 +/− 0.05
7	0.77	0.04	0.77 +/− 0.05	0.77 +/− 0.03	0.78 +/− 0.04	0.77 +/− 0.04
10	0.95	0.03	0.97 +/− 0.02	0.96 +/− 0.02	0.96 +/− 0.03	0.95 +/− 0.03

Finally, we have conducted an additional test for 3, 5, 7, and 10 Red attackers, varying the single shot kill probability (SSKP) of the lethal weapon system from 2% to 30% per round. The results are shown in Fig. 2. The number of Blue killed declines monotonically with the SSKP for all four options. For SSKP values greater than 0.18, the number of Blue killed dropped to zero (within one standard deviation) for seven and fewer attackers, thus rendering these options indistinguishable. Hence, the findings noted in Fig. 1 would not hold for more lethal weapons. This implies that the number of Blue killed depends on the range of the SSKP values used in the sample. If the samples covered only the SSKP values at the minimum or maximum end of the considered interval, the conclusions would differ substantially

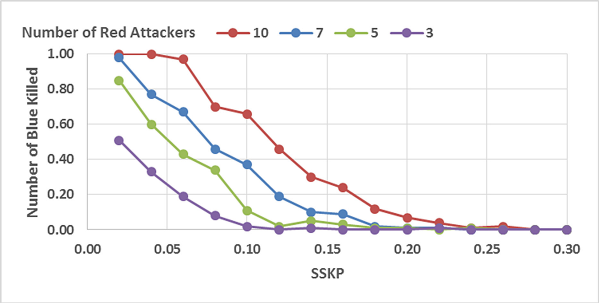


Fig. 2. Dependence of number of Blue killed on the lethal SSKP for 3, 5, 7, and 10 attackers.

The most significant observation to note is that the range of considered parameters needs to be chosen very carefully to cover all feasible options without generating extraneous scenarios. In effect, we would need to run a series of experiments to find the D-optimal design, which contains the best subset of all possible experiments [18], under uncertainty.

4 Summary and Conclusions

In the previous sections, we looked at scenario selection and the resulting comprehensiveness of the representation of the future. First, we examined the basic parameterization required at the operational level to capture the characteristics of individual vignettes and a methodology to determine their likelihood (and duration). We have shown that if there is a historical precedent of a particular type of operation, a probabilistic estimate for the likelihood and duration of a particular vignette can be estimated with no prior knowledge of the sample population. In the cases where vignettes are projecting forward to events that are dissimilar from all prior events, then subject matter expert opinion need to be sought instead to provide some justification for the likelihood of a potential vignette.

Then we addressed the issue of the parameterization of tactical vignettes. A simple scenario looking at naval force protection against asymmetric threats was implemented in an agent-based model called MANA. Selected specifications were varied to determine the dependence of the solution. The results demonstrated that some parameter selection in the scenarios had the potential to significantly influence the findings. It reinforced the requirement to carefully determine the ranges of feasible parameters, especially in the case of complex interactions that are likely to lead to non-linear dependence of the results on the parameters. Design of experiments methods, such as D-optimal designs, could be utilized to select the parameters of interest and the necessary range of values in order to minimize the scenario set used to represent the future solution space.

References

1. Hobson, B.: Obsolescence challenges, Part 3: identifying future capability requirements. Can. Naval Rev. **4**(4), 10–14 (2009)
2. Bowden, F.D.J., Pincombe, B., Williams, P.B.: Feasible scenario spaces: a new way of measuring capability impacts. In: 21st International Congress on Modelling and Simulation, Gold Coast, Australia (2015). www.mssanz.org.au/modsim2015. Accessed 11 Apr 2018
3. Taylor, B., Wood, D.: Guide to capability-based planning, TR-JSA-TP3-2-2004 (2004)
4. Taylor, B.: Toward an Enhanced Capability Based Planning Approach, DRDC-RDDC-2017-D063 (2017)
5. Williams, P.B., Bowden, F.D.J.: Dynamic morphological exploration. In: 22nd National Conference of the Australian Society for Operations Research, Adelaide, Australia (2013). www.asor.org.au/conferences/asor2013. Accessed 11 Apr 2018
6. Chief of Force Development: The Future Security Environment 2013 – 2040, NDID # A-FD-005-001/AF-003 (2014)
7. Chuka, N., Friesen, S.K.: Divining the Force Planning Scenarios: Methodology, Experiences and Lessons from the Canadian Department of National Defence, RTO-MP-SAS-088, RTO System Analysis and Studies Panel Specialists' Meeting, Sweden (2011)
8. Davis, P.K., Bankes, S.C., Egner, M.: Enhancing Strategic Planning with Massive Scenario Generation Theory and Experiments, RAND Corporation (2007). https://www.rand.org/content/dam/rand/pubs/technical_reports/2007/RAND_TR392.pdf. Accessed 18 Apr 2018
9. Wesolkowski, S., Eisler, C.: Capability-based models for force structure computation and evaluation, NATO Workshop on Integrating Modelling & Simulation in the Defence Acquisition Lifecycle and Military Training Curriculum, STO-MP-MSG-126 (2015)
10. Taleb, N.N., The Black Swan: The Impact of Highly Improbable, Random House Trade Paperbacks, 2 edn. (2010)
11. Yao, W.-M., et al.: (Particle Data Group): Statistics. J. Phys. **G33**(1) (2006). http://pdg.lbl.gov/2006/reviews/statrpp.pdf. Accessed 16 Apr 2018
12. Kotz, S., van Dorp, J.R.: Beyond Beta: Other Continuous Families of Distributions with Bounded Support and Applications, pp. 1–28. World Scientific Publishing Co. Pte. Ltd., Singapore (2004). https://www.worldscientific.com/worldscibooks/10.1142/5720#t=toc. Accessed 16 Apr 2018
13. Dobias, P., Eisler, C.: Modeling a naval force protection scenario in MANA. Oper. Res. Manag. Sci. Lett. **1**(1), 2–7 (2017)
14. Lauren, M.K., Stephen, R.T.: Map-aware non-uniform automata (MANA) – a New Zealand approach to scenario modelling. J. Battlefield Technol. **5**, 1–13 (2002)
15. McIntosh, G.C., Galligan, D.P., Anderson, M.A., Lauren, M.K.: MANA (Map Aware Non-Uniform Automata) Version 4 User Manual, DTA TN 2007/3 NR 1465 (2007)
16. Dobias, P., Bouayed, Z., Woodill, G., Bassindale, S.: Optimal number of non-lethal launchers study - Nickel Abeyance II, DRDC CORA TR 2006-18 (2016)
17. Shapiro, S.S., Wilk, M.B.: Analysis of variance test for normality. Biometrika **52**(3), 591–611 (1965). http://www.jstor.org/stable/2333709. Accessed 16 Apr 2018
18. Triefenbach, F.: Design of Experiments: The D-Optimal Approach and Its Implementation as a Computer Algorithm, Thesis, Umea University (2008). http://www8.cs.umu.se/education/-examina/Rapporter/FabianTtriefenbach.pdf. Accessed 20 Apr 2018

Toward a Quantitative Approach to Data Gathering and Analysis for Nuclear Deterrence Policy

Laura Epifanovskaya[1(✉)], Kiran Lakkaraju[2], Josh Letchford[1], Mallory Stites[2], and Jason Reinhardt[1]

[1] Sandia National Labs, Livermore, CA, USA
lwedwar@sandia.gov
[2] Sandia National Labs, Albuquerque, NM, USA

1 Introduction

The doctrine of nuclear deterrence and a belief in its importance underpins many aspects of United States policy; it informs strategic force structures within the military, incentivizes multi-billion-dollar weapon-modernization programs within the Department of Energy, and impacts international alliances with the 29 member states of the North Atlantic Treaty Organization (NATO). The doctrine originally evolved under the stewardship of some of the most impressive minds of the twentieth century, including the physicist and H-bomb designer Herman Kahn, the Nobel Prize-winning economist Thomas Schelling, and the preeminent political scientist and diplomat Henry Kissinger. Following the collapse of the primary U.S. nuclear adversary, the Soviet Union, however, the amount of scholarship dedicated to studying this important topic declined markedly. Since then, the world has undergone many rapid transformations. Important new technologies, including the internet and artificial intelligence, have permeated all elements of society and transformed the nature of everything from manufacturing to business to combat. As the internet has connected people, business and investment have connected nations and state economies have become deeply intertwined. Because of this, the nature of international relationships itself has changed.

This year, the United States completed another Nuclear Posture Review. As national leadership contemplates the role of nuclear weapons as part of our international political and military doctrine, it must now consider the changed nature of the world and the ramifications of these changes. Deterrence has become "cross-domain", where conflict escalation dynamics that might lead to nuclear use can also include cyber, economic, and space elements, interplaying with each other in potentially complex and unanticipated ways. As the nature of strategic

Sandia National Laboratories is a multimission laboratory managed and operated by National Technology & Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell International Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525.

A. J. Morales et al. (Eds.): ICCS 2018, SPCOM, pp. 245–252, 2018.
https://doi.org/10.1007/978-3-319-96661-8_26

deterrence has changed, so must the approach to studying and developing deterrence doctrine. The technologies that have transformed society can also help to simplify highly complex problems and, used properly, bolster our understanding of the complicated interplay of forces that, whether unintentionally or through deliberate policy action, we harness to shield the nation from an act of nuclear aggression by an adversary state. Our team of researchers is developing a capability that leverages the remarkable power of the internet and advancements in machine learning to study nuclear deterrence in a new, data-driven way, using a war game conducted online that explicitly ties back to academic conflict research.

Soon after the advent of the Cold War and building on work begun in RAND and academia, military planners adopted war games as an approach to understanding conflict escalation dynamics that might lead to the use of a nuclear weapon [1]. Although criticized for lacking in scientific rigor in their design, execution, and analysis [2], war games provide an approach to prepare for events that have little historical precedent, such as nuclear conflict. These games can incorporate both computer models and human decision makers. Scenario-based, they can be useful for exploring available options within a given hypothetical event, including options that might not occur to an analyst outside of the game framework [3]. While such exploration is valuable, quantitative post-game analysis is made difficult by the typical game format, which limits the number of plays, the number of players involved, and the ability to capture data in a reproducible way. Specifically, the circumscribed number of turns within a war game yields small data sets not suitable for statistical analysis; the small player base leads to outcomes biased by the individual expertise and personalities of the players themselves; and data may be captured imperfectly in the form of written narratives by human rapporteurs. Because of these limitations, results are typically not analyzed across multiple wargames, although recently there has been some effort to comprehensively compare decisions made within wargames to decisions that the public would make regarding the willingness to use nuclear weapons. In general, war gaming is more useful in illuminating tactical approaches within a given escalation scenario than it is in informing broader strategic decision-making and the effects of particular policies on the dynamics of conflict. As such, a strategic nuclear deterrence architecture is difficult to examine comprehensively using a typical war gaming approach.

For a quantitative approach to analyzing the effects of an overall doctrine on conflict outcomes, we turn to the academic literature. Professors in the fields of policy and international relations study the effects of particular policies and conditions on the likelihood of conflict between states using the Militarized Interstate Disputes (MIDS) data from the Correlates of War Project (correlatesofwar.org), which comprises a database of data on conflict, political alliances, military and economic strength, trade, and other indicators going back around 150 years. Extensive statistical analysis, typically linear regression analysis, has been performed on this data in publications spanning multiple decades (and drawing mixed conclusions about correlations between and among the variables). While this approach to analyzing war is more quantitative (and thus reproducible) than war gaming, it also suffers from drawbacks. First, though data capturing the same parameters across multiple conflicts is available, the MIDS data does not include all the metadata contributing to a given decision leading up to a conflict, and thus may leave crucial information out of the analysis. Second, the data is encoded in dyad-years, which leads to two distinct problems: long conflicts (the Korean War is an example) make outsized contributions to the sample; and the analysis only accounts for relationships between pairs of states (dyads), when in fact conflicts and the decisions surrounding them often involve networks of multiple states. Third, when studying nuclear escalation dynamics, data from purely conventional wars are inadequate to the job. Finally, more generally, this type of analysis is simply too broad and high-level to apply to strategic deterrence and to study potential escalation dynamics within a given deterrence framework.

In order to comprehensively and quantitatively study nuclear deterrence and conflict escalation that might lead to the use of a nuclear weapon, we have formulated an approach that combines the strengths of a war game (including the ability to explore a broad decision-space, record interactions among multiple players, and examine decisions in real time, capturing the available metadata) with the reproducible, mathematical methodology of the conflict literature. We propose to use an online computer war game that enables us to explicitly track and store data in a way analogous to the Correlates of War Project, potentially allowing us to compare analysis of war game data directly to academic research. In addition to MIDS variables, however, this would create the opportunity to track other variables of interest at our discretion, and to record metadata in the form of player conversations through a built-in chat feature. Using the online format, we would have the ability to expand or contract our player base as desired, broadening the player demographic to gather large datasets across a diverse knowledge and experience base for purposes of statistical analysis and to explore the full parameter space, while also maintaining the ability to limit particular game instantiations to one or more select groups of players, such as subject matter experts in the area of nuclear policy or military strategy.

Such a game might comprise a virtual world of states that occupy territory on a virtual landscape, control resources, and have the ability to trade with each other or enter into conflict with each other, both conventional and nuclear. Trade and conflict would be incentivized by an uneven distribution of resources among states. Players would have some goals, or "victory conditions", in common, but individual players would also have unique, hidden victory conditions, meant to capture the unique motivations of different states in the real world. The game would be designed to explicitly capture data on MIDS variables such as trade, alliance, military and economic strength, and geographical contiguity. In addition, we would design features into the game to capture important elements of deterrence such as signaling interactions between states, as well as allowing for uncertainty in both signaling intentions to other players and in success, should one player choose to attack another. We plan to use such a game to explore specific research questions, while at the same time populating a database with MIDS-like data that can be analyzed retroactively using statistical methods, including modern methods from the discipline of data analytics, and even compared to real-world data from the Correlates of War Project. The database would be made freely available to the public, for use by academics and analysts performing conflict research.

2 Proof of Concept

Before embarking on a new war game design, and in order to demonstrate the viability of our proposed approach, we performed some preliminary analysis. There were three primary questions that we intended to answer before proceeding:

– What do experts on deterrence and nuclear policy think of our approach? Will the results of our proposed war game be valuable to them?
– Is there good reason to expect that data from online games is valuable in evaluating real-life problems?
– Can we use data that we already have from an online game, perform some simple analysis, and draw preliminary conclusions about the viability of online game data in general to answer the questions that we're interested in answering?

To answer the first question, we talked to multiple experts in the areas of international relations, deterrence, and nuclear policy. With one exception, we were encouraged by all to pursue the development of the war game as a way of supplementing existing data and methods of studying the deterrence problem. The dissenting expert considers clarity in the deterrence problem to be antithetical to deterrence itself; essentially, the more ambiguity and uncertainty that exists concerning states' willingness and intent to use nuclear weapons, the better nuclear deterrence works. The others opined that clarity in signaling intentions and objectives only helps to strengthen a deterrence posture; that data relevant to the deterrence problem are needed and would be valuable to academics; and that data from a war game would be a useful substitute for real-world conflict data.

To address the second question, we looked at academic research that explores the relationships between online and real-world behaviors, and thus the usefulness of virtual worlds in studying real-world problems. Unsurprisingly because of the huge potential for internet-based experiments as a source of data for behavioral experiments, this is an active area of academic inquiry. In one investigation, researchers assessed the Big-5 personality traits (openness to experience, conscientiousness, extraversion, agreeableness, and neuroticism) of 1,040 World of Warcraft Players and examined parallels between these traits and player behavior in the online game [4]. In this work, statistically-significant correlations were found between online behavior and player personality traits. As an example, extraverted players tended to exhibit extravert behaviors in the context of the games, such as frequently collaborating in groups. Another study found real-world patterns in economic behaviors in an online game context [5]. In general, participants in Massively Multiplayer Online Games (MMOGs) such as World of Warcraft tend to demonstrate significant investment in and emotional connection with their online avatars, which become a reflection of the players themselves. Emergent, complex social behaviors are often observed in the game context, such as long-term acts of in-game espionage conducted against virtual adversaries [6].

Finally, having addressed the first two questions, we performed a preliminary proof-of-concept using 739 days of online game data gathered as part of an earlier, unrelated research effort. The game in question is a browser-based game which will be referred to here as Game X, and which is premised on a virtual world in which players can perform several actions that emulate real-world behaviors, such as purchasing things of value, forming alliances and trade relationships, and engaging in conflict (See Appendix A for more details). The purpose of our analysis was to take actions performed over the two years of data gathering and "operationalize" them: map them to variables in the Correlates of War database, and then directly apply the methodology from a single, simple example paper from the academic conflict literature to see how the online results compare to the real-world results from the paper. The paper chosen was Professor Katherine Barbieri's 1996 article Economic Interdependence: A Path to Peace or a Source of Interstate Conflict? [7] Though the methodology employed by the paper has detractors within the international relations community, the purpose of our effort was not to draw conclusions about economic interdependence and conflict; but rather to demonstrate that analysis of online game data can be compared directly to results from academic research.

The Barbieri paper evaluates 14,341 dyad-years (a dyad is a two-state relationship, and a dyad-year reports data for that relationship in a given year) over the period 1870-1938. The primary explanatory variables used in her analysis are salience, an indicator of the importance of the trade relationship to the dyad; symmetry, which attempts to quantify the difference in importance of that relationship between the two partners, and interdependence, which measures salience and symmetry together. The salience and symmetry variables are derived using trade share, a measure of trade between the dyad compared to the overall amount of trade conducted by a partner within the dyad.

$$\text{Trade Share}_i = \frac{DyadicTrade_{ij}}{TotalTrade_i} \tag{1}$$

$$\text{Salience} = \text{Sqrt}(TradeShare_i * TradeShare_j) \tag{2}$$

$$\text{Symmetry} = 1 - |\text{TradeShare}_i - \text{TradeShare}_j| \tag{3}$$

$$\text{Interdependence} = \text{Salience} * \text{Symmetry} \tag{4}$$

Again, the purpose of the analysis here is not to support or critique the author's methodology, but simply to duplicate it and compare the Game X analysis to the results of the article. In addition to these explanatory variables, several control variables were used in Barbieri's regression analysis, such as contiguity, alliance, and capabilities ratio. Our analysis included a contiguity indicator, in which lack of contiguous properties held by two guilds (the game version of sovereign states) was encoded as "0" and existence of contiguous properties was encoded as "1". Alliance was coded as "1" if neither guild in a dyad had designated the other as a foe, and "0" otherwise. In lieu of capabilities ratio we used combat strength ratio, economic strength ratio, and size ratio, since these indicators were measured and tracked individually within the game context. The capabilities ratio variable from the Correlates of War Project is a relative measure of the Composite Index of National Capabilities (CINC) scores of the states within a dyad. The CINC score is based on the population, urban population, iron and steel production, energy consumption, military personnel, and military expenditure of a state.

The analysis presented here used a linear mixed effects regression model, run using the *lme4* package for R. The results are best compared to Barbieri's Full Model of Militarized Disputes, which included all of the explanatory variables in the model, and modeled all dyadic disputes, not just major wars. Our analysis similarly includes all disputes between pairs of guilds.

Table 1. Comparison of Game X linear regression coefficients to those from Dr. Katherine Barbieri's 1996 paper *Economic Interdependence: A Path to Peace or a Source of Interstate Conflict?*

Variable	Game X			Barbieri		
	β	SE	p	β	SE	p
Salience	463.27	131.95	<0.001	-22.64	6.69	≤ 0.01
Symmetry	16.01	4.33	<0.001	-4.46	0.80	≤ 0.01
Interdependence	-490.88	140.78	<0.001	26.60	7.28	≤ 0.01

As seen in Table 1, the correlations between the explanatory variables and the dependent conflict variable reported in the Barbieri paper do not agree with those from the Game X analysis. A large, positive value of both salience and symmetry from the Game X analysis indicate that the more important trade relationships are between guilds, and the more symmetric they are between the trading partners, the more likely those guilds are to enter into conflict with each other (although the mathematical definition of these variables within the context of the article may not be the ideal way to capture these two quantities). Similarly to the Barbieri paper, interdependence reflects an opposite correlation to conflict to that of both of its composite variables (unrealistic but logical given the way that the variable is defined mathematically). As in Barbieri's paper, salience is a stronger indicator than symmetry of conflict likelihood.

From this analysis, it is unclear why opposite conclusions emerge about the relationships between the dependent and independent variables from Game X and Barbieri's work. She includes the joint democracy control variable in the regression, which the Game X data does not account for, but which is not a strong predictor of conflict in her model. The Barbieri paper is not explicit about the regression methodology used, further complicating the comparison.

3 Conclusions

In the 20 intervening years between the publication of Economic Interdependence: A Path to Peace or a Source of Interstate Conflict?, much work has been published in academia on the relationship between economic interdependence and conflict disputing the article's methodology and results and suggesting that the relationship may not be as simple as a negative or positive correlation between trade volumes and military disputes. If we are to continue to explore correlations between conflict and MIDs variables in future work, we would like to redefine the explanatory variables from the paper, include additional variables that capture important interstate relationships, and improve upon the simple linear regression techniques used to identify correlations. However, the purpose of this study was not to analyze the relationship between trade and conflict, but to demonstrate that online game data can be operationalized in such a way as to make direct comparisons between it and data from academic research. This we have successfully done, although if we want to make future direct comparisons between academic literature and online research, we will need to explore ways to quantify all of the variables, such as joint democracy, that are included in the academic analysis. This is something to think about explicitly as we move forward with our online game design. A game that explores nuclear deterrence will need to explicitly capture nuclear conflicts and nuclear capabilities, and will preferably also capture variables that are highly relevant to interstate relationships in the modern age but are not explicitly included in Correlates of War, such as foreign direct investment, which today constitutes a much larger percentage of interstate economic interactions than trade [8]. Ideally, variables such as capability ratio would also capture important modern capabilities such as competency in cyberspace.

The time for renewed and energetic study of strategic nuclear deterrence is now, and our team of researchers from academia and the national laboratories is applying itself to that study with a full complement of modern methods. By leveraging the lessons of several decades of war gaming, applying the quantitative constructs of academia, and harnessing the data-gathering power of the internet coupled with state-of-the-art techniques in artificial intelligence, we hope to provide national decision makers with an empirical tool to inform nuclear deterrence policy deliberations, including future Nuclear Posture Reviews.

A Appendix: Description of Game X

Game X is a browser-based exploration game which has players acting as adventurers owning a vehicle and traveling a fictional game world. There is no winning in Game X, rather players freely explore the game world and can mine resources, trade, and conduct war. There is a concept of money within Game X, which we refer to as marks. To buy vehicles and travel in the game world players must gather marks. There is also a vibrant market-based economy within Game X.

Players can communicate with each other through in-game personal messages, public forum posts and chat rooms. Players can also denote other players as friends or as hostiles. They can take different actions, such as: move vehicle, mine resources, buy or sell resources, build vehicles, products, and factory outlets. Players can use resources to build factory outlets and create products that can be sold to other players.

Players can also engage in combats with non player characters, other players, and even market centers and factory outlets. They can modify their vehicles to include new weaponry and defensive elements. Players have "skills" that can impact their ability to attack/defend.

A.1 Groups in Game X

There are four types of groups a player may belong to. Table 2 summarizes the properties of these. Of these, guilds are the most pertinent to our study.

Table 2. Summary of properties of the groups in Game X

	Nation	Agency	Guild	Race
Number:	Fixed, 3	Fixed, 2	Dynamic, Many	Fixed
Membership type	Open	Open (req's)	Closed	Open
Modifiable	Yes	Yes	Yes	No

Guilds allow members to cooperate to gain physical and economic control of the game world. Guilds are comprised of a leader and a board who form policy and make decisions that impact the entire guild members. Guilds can be created by any player once they have met the experience and financial requirements. Guilds are *closed* – players must submit an application and can be denied of membership.

References

1. Pauly, R.: Nuclear Weapons in Wargames: Testing Traditions and Taboos. SSRN Scholarly Paper ID 2917628, Social Science Research Network, Rochester (2017). https://papers.ssrn.com/abstract=2917628
2. Brewer, G.D., Shubik, M.: The war game: a critique of military problem solving. Harvard University Press, Cambridge (1979). https://catalog.hathitrust.org/Record/000177835
3. Carter, A.B., Steinbruner, J.D., Zraket, C.A.: Managing Nuclear Operations. Brookings Institution (1987)
4. Yee, N., Ducheneaut, N., Nelson, L., LIkarish, P.: Proceedings of the CHI 2011 (2011). http://www.nickyee.com/pubs/Yee%20(CHI%202011)%20-%20Personality%20in%20WoW.pdf
5. Castronova, E., Williams, D., Shen, C., Ratan, R., Xiong, L., Huang, Y., Keegan, B.: Introverted elves & conscientious gnomes: the expression of personality in world of warcraft. New Media Soc. **11**(5), 685 (2009)
6. Francis, T., PC Gamer Magazine, p. 90 (2006)
7. Barbieri, K.: Economic interdependence: a path to peace or a source of interstate conflict? J. Peace Res. **33**(1), 29 (1996). http://www.jstor.org/stable/425132
8. Blackwill, R.D., Harris, J.M.: War by Other Means: Geoeconomics and Statecraft. Harvard University Press (2016). http://www.jstor.org/stable/j.ctt1c84cr7

CityScope: A Data-Driven Interactive Simulation Tool for Urban Design. Use Case Volpe

Luis Alonso(✉), Yan Ryan Zhang, Arnaud Grignard, Ariel Noyman, Yasushi Sakai, Markus ElKatsha, Ronan Doorley, and Kent Larson

MIT Media Lab - City Science Group, Cambridge, USA
alonsolp@media.mit.edu

Abstract. MIT City Science Group (CS) studies the interaction of social, economic and physical characteristics of urban areas to understand how people use and experience cities with the goal of improving urban design practices to facilitate consensus between stakeholders. Long-established processes of engagement around urban transformation have been reliant on visual communication and complex negotiation to facilitate coordination between stakeholders, including community members, administrative bodies and technical professionals. City Science group proposes a novel methodology of interaction and collaboration called CityScope, a data-driven platform that simulates the impacts of interventions on urban ecosystems prior to detail-design and execution. As stakeholders collectively interact with the platform and understand the impact of proposed interventions in real-time, consensus building and optimization of goals can be achieved. In this article, we outline the methodology behind the basic analysis and visualization elements of the tool and the tangible user interface, to demonstrate an alternate solution to urban design strategies as applied to the Volpe Site case study in Kendall Square, Cambridge, MA.

1 Introduction

In an era when more than 50% of the world's population lives in cites and more than 80% of the global GDP is generated in urban centers, cities are becoming increasingly important for societal well being. Global populations are urbanizing at unprecedented rates and scales, mandating the need for urban planning and design practices that are scalable, streamlining much of the decision-making processes. Long-established urban planning practices, which are heavily reliant on institutional capacity and complex procedures, are falling short of the demands, resulting in single-use, auto-centric development patterns, and environmental degradation. However, collaboration tools that are widely accessible and provide evidence-based feedback, can greatly improve urban management processes.

Design and analytical tools such as CAD and GIS were developed for single-user professionals, and as such, feature limited capabilities for interaction

A. J. Morales et al. (Eds.): ICCS 2018, SPCOM, pp. 253–261, 2018.
https://doi.org/10.1007/978-3-319-96661-8_27

(mainly mouse, keyboard and screen)[16]. Subsequent design and technology pioneers developed alternative interfaces that facilitate a more collaborative process of urban design, aided by computation and data analytics. In the early 2000s, a series of projects developed Tangible-User Interfaces (TUIs) for real-time design and collaboration. The Augmented Urban Planning Workbench [8], The Clay Table [9], The I/O bulb and the Luminous Room [19] mixed traditional and computational design processes using TUIs augmented by computational analysis. Larson's 'Louis I. Kahn - Unbuilt Ruins' exhibition [12] utilized location tagged-building objects to allow users to localize themselves and navigate through a virtual built environment. These experiments all provided TUI innovations and proof-of-concept which eventually enabled the development of a more scalable, versatile platform for collaborative evidence-based urban planning.

In the past few years, the MIT City Science Group has been developing CityScope, a TUI platform that aims to reduce much of the coordination and collaboration challenges associated with traditional planning methodologies. CityScope facilitates consensus building through a participatory process, which allows community members, local authorities and design professionals to understand challenges, explore alternatives and receive real-time feedback in response to proposals and/or interventions. The system provides a framework to discuss the potential impact of proposals and policy interventions. It has three primary objectives: (i) to visualize complex urban data and the inter-relationships between urban systems, (ii) to simulate the impact of urban interventions and (iii) to support decision making in a dynamic, iterative, evidence-based process using a tangible interface.

The paper is organized as follows. Section 2 describes the main urban performance indicators incorporated into the CityScope platform. Section 3 describes a use case for the CityScope platform on the Volpe site. Finally Sect. 4 concludes the paper with a discussion of the strengths of this approach and the opportunities for further development.

2 Urban Performance Indicators

As shown in Fig. 1 CityScope includes three core modules: (i) a physical 3D model, (ii) a tangible interactive interface for interventions and (iii) a computational analysis unit with a feedback system. In this section we focus on describing the urban performance indicators shown in Fig. 3. The computational analysis unit uses a number of tools to calculate a range of performance metrics as outlined in this section and produces feedback in the form of both spatial metrics and abstract statistics, which have been implemented using different software tools including Rhino [1], GAMA platform [6] and Unity [16].

Examples of spatial feedback include, but are not limited to, agent-based models for transposition and mobility information, visual graphs and heat-maps for proximity of residences to workplaces, and collision potential for interaction. The spatial feedback is mapped onto the 3D model using an array of vertical projections. The abstract metrics include global metrics of urban performance in

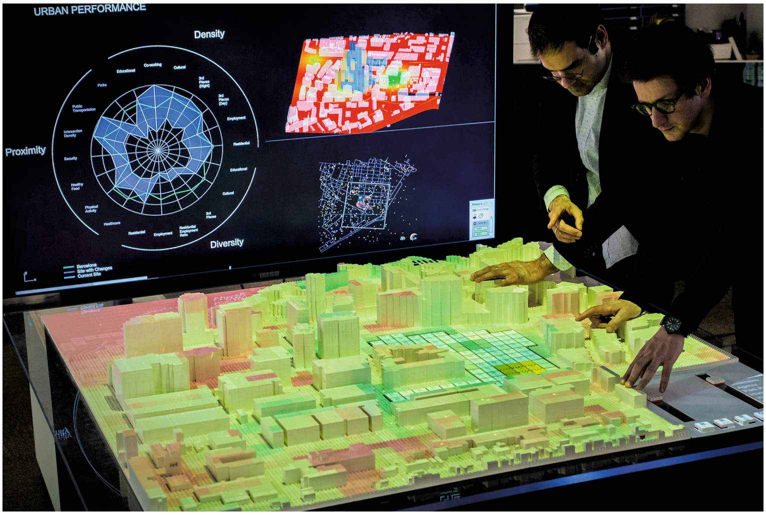


Fig. 1. CityScope Volpe: a data-driven interactive simulation tool for urban design.

relation to density, proximity, diversity, energy usage and innovation potential. These metrics are displayed using visual tools such as radar plots and bar plots and are displayed on a vertical screen as shown in Figs. 1 and 2. The combination of the tangible user interface and the real-time feedback systems allows users and stakeholders to collaborate more effectively by collectively performing rapid experimentation and receiving data-driven feedback.

As Jacobs describes in [3], fundamental correlations exists between urban form and urban performance - including quality of life, vibrancy, and safety. In this work, urban performance is analyzed and described in terms of social, economic and physical attributes or metrics, which are categorized under the headings of Density, Diversity, Proximity, Mobility, Building Energy and Innovation Potential [10]. This research is focused on understanding how these urban indicators interact to support well functioning urban ecosystems. The rest of this

Fig. 2. Abstraction process from the actual urban area into an interactive LEGO model which interfaces with mathematical models and simulations.

section describes the calculation of some of the key indicators used for real-time feedback in CityScope.

2.1 Density

Cities can be understood as the right ratio between housing, commercial space, retail venues, cultural institutions, and other amenities. This ratio must be calibrated to ensure positive outcomes resulting from increased density (greater exchange of ideas, higher wages, more employment options, better amenities, etc.), but with fewer negative outcomes such as traffic congestion, stress, loss of contact with nature, pollution, etc. [7]. The density of each building type or population (eg. amenities, residences, employment) can be computed as follows.

$$Density^m = \frac{|M|}{A}$$

M = set of occurrences of m in district
A = area of district

2.2 Diversity

Well balanced urban ecosystems require a diversity of social, economic and spatial characteristics to ensure high levels of urban performance and resilience. The Diversity index is assessed using the Shannon-Weaver formula or [11] Balanced Ecosystem index: Quantitative measurement that reflects how many different types (species) there are in a dataset (community/ecosystem), and simultaneously takes into account how evenly the basic entities (individuals) are distributed among those types.

$$H = -\sum_{i=1}^{S} p_i ln p_i$$

H = Shannon's diversity index
S = total number of species in the community (richness)
p_i = proportion of the population made up of species i

2.3 Proximity

Urban proximity is used to assess the accessibility and closeness indexes between social, economic and physical activities and characteristics of urban environments [18]. Higher performing future cities will likely resemble networks of compact urban districts where resources and amenities of daily life are in close proximity to one another in walkable, vibrant communities [20].

CityScope can be used to test such adjacency scenarios, with near real-time feedback as parameters are changed through iterations. Metrics include walkable access to parks, housing, jobs, mass transit, schools, amenities, and other parameters, depending on the context. For a given 'goal object', a proximity metric for a district can be calculated as follows.

$$Closeness\,to\,SM = (\sum_{j \epsilon B_i} dist(j, closest(j, SM)))^{-1}$$

B_i = Set of blocks in district i
SM = Set of "goal" objects (Parks, Residential, Office ...)
$dist(a; b)$ = Geographical distance between two elements' centroids a and b
$closest(j; Y)$ = Geographically closest element in set Y from block j's centroid

2.4 Mobility

The use of Agent Based Modeling (ABM) in CityScope allows for the rapid exploration of mobility alternatives to private automobiles, including shuttle-on-demand systems, bike-sharing, on-demand autonomous vehicles, and other mobility strategies that enable an increase in density without traffic congestion and land-uses for vehicle storage. A model which simulates an artificial micro-environment and reacts to changes in land-use and mobility modes is used to explore mobility patterns. The model is described more in detail in [4]. The impacts of urban mobility can be quantified in a number of ways. For example, given a set of mobility patterns estimated from an ABM or an econometric approach, the total mobility energy can be calculated as follows:

$$E_m = \sum_{m \in M} \sum_{l \in L} v_l^m s_l e^m$$

E_m = Total mobility energy
M, L = The set of all transport modes and road network links respectively
v_l^m = The number of vehicles of mode m that travel on link l
s_l = The length of link l
e_m = The energy usage per unit distance traveled by mode m

2.5 Building Energy

Energy Consumption is calculated for structures within the the area of investigation using data about building envelope performance, building orientation, estimated embodied energy, and land-use information. CityScope aims to predict how overall energy consumption of the buildings will change given different urban configurations at preselected times of the year. The model uses a combination of building use, height, and proximity to adjacent structures (to account for solar gain) to determine the overall energy efficiency of the buildings within the area under consideration.

The index of energy efficiency is then found by dividing the total energy metrics of the buildings by the total energy input. [2,17]

$$Energy\,Efficiency = \frac{\sum(Useful\ Energy\ Output)}{\sum(Energy\ Input)} \times 100\%$$

2.6 Innovation Potential

The ABM models enable the study of innovation potential in the city by creating a spatial graph where the agents are the node of the graph and where the edges are formed when an agent is close to one another. This graph is used to compute metrics such as density, degree distributions, correlations and centrality to indicate the likelihood of innovation in the city. [10] It is proposed that the total innovation potential in a district is related to the total collision potential between agents of different demographic groups. This collision potential can be defined as follows:

$$C_{A,B} \propto \sum_{a \in A} \sum_{b \in B} \delta_{a,b} \quad \forall A, B \in D$$

where

$$\delta_{a,b} = \begin{cases} 1, & \text{if } s^t_{ab} < s_{max} \text{ for any time, } t \\ 0, & \text{otherwise} \end{cases}$$

$C_{A,B}$ = The collision potential between demographic groups A and B.
D = The set of all demographic groups under study
s^t_{ab} = The distance between agent a and agent b at time t
s_{max} = The maximum separation distance for interaction potential

3 MIT- Kendall Square Case Study

Over the past 20-years, Kendall Square in Cambridge, MA has undergone an extreme transformation as a result of the emerging biotech industry and knowledge-based economy,largely driven by the proximity to MIT and Harvard. Although workers in and around Kendall Square could benefit from living close-by, a limited housing stock and high land values make most residential opportunities in the area out of reach for young professionals, families, and seniors. With a residential density of only 3,000/km^2, most people who work in Kendall Square commute long distances each day at the cost of energy and time. Affordable housing incentives for developers are inadequately promoting the necessary range of housing options, and the zoning ordinance has overly restrictive land-use requirements [14]. Parallel to low residential density there is great scarcity of services, amenities and 3rd-places, such as coffee shops, restaurants and bars, which tend to indicate the urban vitality of a place [13].

CityScope Volpe platform was built to investigate the redevelopment of the 14-acre D.O.T facility recently purchased by MIT. The physical 3D model is built using LEGO and covers a region of 1 km × 1 km at a scale of 1:762 m; one 4 × 4 LEGO tile represents a 26.7 × 26.7 m area. The abstraction provided by the model helps avoid high levels of detailed building-form and urban-design, allowing users to focus on the land-use and planning aspects of the urban environment. In addition, LEGO models provide a tangible interface easy to maintain, redesign and manipulate.

The TUI interactive section is integrated into the modeled surroundings to represent the area under development, Volpe site in this case as shown in Fig. 2. This intervention region utilizes CityMatrix [21], a CityScope tool designed to allow users to modify land-use and urban density by manipulating predefined physical pieces of LEGO tiles. Interchangeable tiles can represent different land-uses ranging from roads, parks, amenities, residential or office buildings. CityMatrix includes a physical UI components with sliders and toggles to change the building height and switch between alternative mobility modes. When users add or move these pieces, sliders, and toggles, they reshape the design of the urban area under study. These modifications are recorded and passed to the computational analysis unit in order to provide real-time feedback to the user.

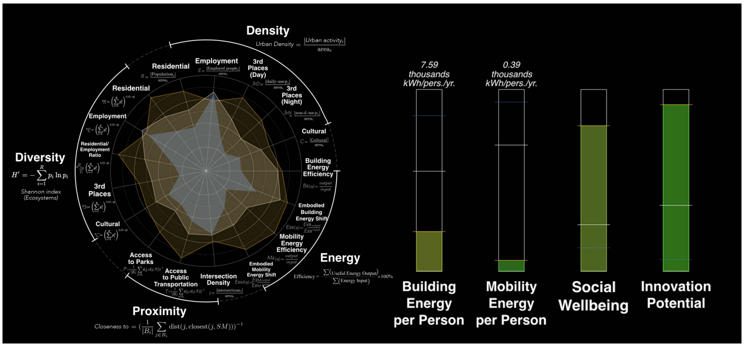


Fig. 3. Urban performances indicators of the district updates in real-time: density, diversity, proximity, mobility energy per person, social wellbeing (walkability, access to parks, etc.), and innovation potential (probability of creative collisions).

4 Conclusion

CityScope platform developed by City Science group at the MIT Media Lab leverages data-analytics, computation, visualization methods and an intuitive interface, to offer advantages over traditional planning and urban design practices. It allows for the rapid assessment of urban interventions which incorporates analytical and visualization components to enable dynamic or iterative, evidence-based decision making between traditionally siloed stakeholders ranging from community members and government officials, to domain experts and technicians. These intuitive data-visualizations help to simplify complex urban systems for users of varying disciplines and levels of expertise. The combination of a tangible user interface and real-time feedback facilitate consensus-building through collaborative experimentation allowing multiple stakeholders to address a wide range of interests simultaneously. Finally, the analysis and models which underly the feedback allow the planning process to become more transparent, data-driven and evidence-based.

Since early 2016, the Volpe platform has been used for numerous daily demonstrations, workshops and interactions with both the MIT community and decision makers. The interactions are being used to understand user reactions to the interface, as well as the underlying analytics behind the results. Thus far, the system has been adjusted on an ongoing basis to improve user experience. The next step of the research will be to conduct controlled workshops to verify the UI as well as the analytical components of the tool. In addition to the Volpe table, the CityScope platform has been developed and refined through a series of deployments in cities around the world that make up the growing City Science Network [5,15]. Cities include Shanghai, Helsinki, Hamburg and Andorra all of who are working on different urban challenges. The CityScope platform continues to evolve in order to respond to the biggest challenges faced in varied urban environments of the global cities.

References

1. Day, M.: Rhino grasshopper (2009). Accessed 30 June 2014
2. Ferrão, P., Fernández, J.E.: Sustainable Urban Metabolism. MIT Press, Cambridge (2013)
3. Fuller, M., Moore, R.: The Death and Life of Great American Cities. Macat Library (2017)
4. Grignard, A., et al.: Simulate the impact of the new mobility modes in a city using ABM. In: ICCS 2018 (To be published) (2018)
5. Grignard, A., Macià, N., Alonso Pastor, L., Noyman, A., Zhang, Y., Larson, K.: Cityscope andorra: a multi-level interactive and tangible agent-based visualization (extended abstract). In: AAMAS 2018 (To be Published) (2018)
6. Grignard, A., Taillandier, P., Gaudou, B., Vo, D.A., Huynh, N.Q., Drogoul, A.: GAMA 1.6: advancing the art of complex agent-based modeling and simulation. In: International Conference on Principles and Practice of Multi-Agent Systems, pp. 117–131. Springer (2013)
7. Hawley, A.H.: Population density and the city. Demography **9**(4), 521–529 (1972)
8. Ishii, H., et al.: Augmented urban planning workbench: overlaying drawings, physical models and digital simulation. In: Proceedings of the 1st International Symposium on Mixed and Augmented Reality, p. 203. IEEE Computer Society (2002)
9. Ishii, H., Ratti, C., Piper, B., Wang, Y., Biderman, A., Ben-Joseph, E.: Bringing clay and sand into digital design–continuous tangible user interfaces. BT Technol. J. **22**(4), 287–299 (2004)
10. Katz, B., Wagner, J.: The Rise of Innovation Districts: A New Geography of Innovation in America. Brookings Institution, Washington (2014)
11. Kemeny, T.: Immigrant diversity and economic development in cities: a critical review (2013)
12. Larson, K.: Louis I. Kahn: Unbuilt Masterworks. Monacelli Press, New York (2000)
13. Montgomery, J.: Editorial Urban Vitality and the Culture of Cities (1995)
14. Noyman, A.: POWERSTRUCTURES: the form of urban regulations. Master's thesis, Massachusetts Institute of Technology (2015)
15. Noyman, A., Holtz, T., Kröger, J., Noennig, J.R., Larson, K.: Finding places: HCI platform for public participation in refugees' accommodation process. Procedia Comput. Sci. **112**, 2463–2472 (2017)

16. Noyman, A., Sakai, Y., Larson, K.: Cityscopear: urban design and crowdsourced engagement platform. In: CHI 2018 CHI Conference on Human Factors in Computing Systems (To be Published) (2018)
17. Patterson, M.G.: What is energy efficiency?: concepts, indicators and methodological issues. Energy Policy **24**(5), 377–390 (1996)
18. Sevtsuk, A., Kalvo, R., Ekmekci, O.: Pedestrian accessibility in grid layouts: the role of block, plot and street dimensions. Urban Morphol. **20**(2), 89–106 (2016)
19. Underkoffler, J.: A view from the luminous room. Pers. Technol. **1**(2), 49–59 (1997)
20. Westerink, J., Haase, D., Bauer, A., Ravetz, J., Jarrige, F., Aalbers, C.B.: Dealing with sustainability trade-offs of the compact city in peri-urban planning across European city regions. Eur. Plann. Stud. **21**(4), 473–497 (2013)
21. Zhang, Y.: CityMatrix - An Urban Decision Support System Augmented by Artificial Intelligence. Master's thesis, Massachusetts Institute of Technology (2017)

Transitivity vs Preferential Attachment: Determining the Driving Force Behind the Evolution of Scientific Co-Authorship Networks

Masaaki Inoue[1,2(✉)], Thong Pham[2], and Hidetoshi Shimodaira[1,2]

[1] Kyoto University, Kyoto, Japan
minoue@sys.i.kyoto-u.ac.jp
[2] RIKEN AIP, Tokyo, Japan

Abstract. We propose a method for the non-parametric joint estimation of preferential attachment and transitivity in complex networks, as opposite to conventional methods that either estimate one mechanism in isolation or jointly estimate both assuming some functional forms. We apply our method to three scientific co-authorship networks between scholars in the complex network field, physicists in high-energy physics, and authors in the Strategic Management Journal. The non-parametric method revealed complex trends of preferential attachment and transitivity that would be unavailable under conventional parametric approaches. In all networks, having one common collaborator with another scientist increases at least five times the chance that one will collaborate with that scientist. Finally, by quantifying the contribution of each mechanism, we found that while transitivity dominates preferential attachment in the high-energy physics network, preferential attachment is the main driving force behind the evolutions of the remaining two networks.

Keywords: Preferential attachment · Clustering coefficient
Rich-get-richer · Transitivity · Scientific co-authorship networks
Collaboration networks

1 Introduction

Cooperation is among the most fundamental behaviors of living creatures [1]. Animals cooperate in various activities: from hunting and forming territories to grooming and child raising [2]. Humans are the experts of cooperation. From cooperation between states [3], companies [4] to cooperation between individuals [5], it is the bedrock of our society.

As a form of human cooperation, scientific collaboration is the backbone of the scientific world. In this process, scientists share their ideas, their time, and their skills with each other in order to push the boundary of knowledge. Since the start of the twentieth century, the number of scientific articles with

A. J. Morales et al. (Eds.): ICCS 2018, SPCOM, pp. 262–271, 2018.
https://doi.org/10.1007/978-3-319-96661-8_28

more than one author has grown to more than three times the number of single-author articles [6]. There is accumulating evidence that articles resulted from collaborations are cited more frequently than non-collaborated ones [6,7]. Since the number of citations is the main metric of scientific impact [8], collaborations thus lead to high impact research. Therefore, understanding the evolution of a scientific co-authorship network, in particular, understanding how new collaborations are fostered, is significantly important for policy makers, funding agencies, university managers as well as each scientist.

To understand the evolution of a scientific co-authorship network, it is beneficial to consider this evolution in a larger context of the evolution of a complex network. Two defining characteristics of an evolving complex network are the heavy-tail of the degree distribution and the high value of the clustering coefficient [9]; both are often represented at the same time in scientific co-authorship networks [10,11].

On the one hand, to explain the heavy tail property, complex network studies have proposed the preferential attachment (PA) mechanism where the probability that a node with degree k receives a new link is proportional to the PA function A_k [12,13]. When A_k is an increasing function on average, nodes with higher numbers of links will receive more new links, and thus hubs are formed and the heavy-tail degree distribution emerges.

On the other hand, one of the most simple mechanisms to explain the high value of the clustering coefficient is transitivity where the probability that a pair of nodes with b common neighbors receives a new link is proportional to the transitivity function B_b [14]. When B_b is an increasing function on average, more triangles are formed between sets of three nodes, and this leads to an increase in the clustering coefficient.

Existing approaches either estimate one mechanism in isolation [13–15] or estimate jointly the two mechanisms assuming some parametric forms for A_k and B_b [16,17]. On the one hand, estimating either mechanism in isolation often leads to poor fit, since many real-world networks exhibit simultaneously heavy-tail degree distribution and high-clustering. On the other hand, it is difficult to justify a particular choice of functional forms used in parametric estimation methods. A non-parametric estimation method would allow the functional forms to be learnt from the observed data.

Estimating A_k and B_b is the first step towards answering the question of what matters more in the evolution of a complex network: transitivity or PA. While there is some research studying a similar question regarding PA and fitness [18], the question regarding PA and transitivity has curiously remained unexplored, despite its potential to provide deeper understanding on how new cooperation is fostered.

Our contribution is threefold. In our first contribution, we propose a method for the non-parametric joint estimation of the PA function A_k and the transitivity function B_b. In our model, the probability that a new edge emerges between node i and node j at time step t is

$$P_{ij}(t) \propto A_{k_i(t)} A_{k_j(t)} B_{b_{ij}(t)}, \tag{1}$$

where $k_i(t)$ and $k_j(t)$ are the degrees of nodes i and j at time-step t, respectively, and $b_{ij}(t)$ is the number of common neighbors between i and j at that time-step. We estimate A_k and B_b by maximum likelihood estimation. The log-likelihood of the observed data is maximized by a Minorize-Maximization algorithm [19]. This is an iterative algorithm that increases the log-likelihood value at each iteration until convergence. Although our primary targets for estimation are the non-parametric functions A_k and B_b, one can apply the same maximum likelihood estimation method for the power-law parametric model where we assume $A_k = (k+1)^\alpha$ and $B_b = (b+1)^\beta$. The method is discussed more in details in Sect. 2.

We use the co-authorship network between scientists in the complex network field [20] to illustrate the goodness-of-fit of three estimation methods: (a) jointly estimating the two mechanisms, i.e., the proposed method in Eq. (1), (b) estimating the PA function A_k in isolation, i.e., $P_{ij}(t) \propto A_{k_i(t)}A_{k_j(t)}$, and (c) estimating the transitivity function B_b in isolation, i.e., $P_{ij}(t) \propto B_{b_{ij}(t)}$. In each case, after estimation, starting from the same initial snapshot as in the real-world data, we simulated the new edges based on $P_{i,j}(t)$ while we kept the numbers of new nodes and new edges at each time-step exactly the same as in the real-world network. Figure 1 shows the clustering coefficient C in the final snapshot of the simulated networks of each case over 100 replications. Only the proposed method of joint estimation could satisfactorily reproduce the true clustering coefficient C_*. This implies that the true growth mechanism in this network, whatever it might be, is closer to PA + transitivity than either PA or transitivity alone. Using, for example, transitivity alone to explain the growth of the network would lead to over-estimation of the transitivity function B_b, since the edge-increasing effect of the neglected PA mechanism has to be explained by B_b alone. This over-estimation, in turn, would lead to an overly high C as can be seen in Fig. 1.

In our second contribution, we propose a method to quantify the contributions of PA and transitivity mechanisms in the growth process of a network, and thus provide a way to answer the question of what matters more: PA or transitivity. The method is discussed in Sect. 3.

In our third contribution, we analyzed three scientific co-authorship networks between scholars in the complex network field [20], physicists in high-energy physics [21], and authors in the Strategic Management Journal (SMJ) [22]. The results are shown in Sect. 4. As opposite to conventional parametric approaches, the non-parametric approach provides us a chance to investigate finer details of PA and transitivity mechanisms. While A_k in CMP and SMJ is increasing in general, its trend is complex in HEP: it is first increasing, then decreasing, and finally increasing again. On the other hand, while B_b in HEP shows a clearly increasing trend, it is flat or even decreasing in CMP and SMJ when we exclude $b = 0$.

Furthermore, in all networks, having at least one common collaborator with another scientist increases at least five times the chance that one will collaborate with that scientist in the future, which is consistent with previous studies [14].

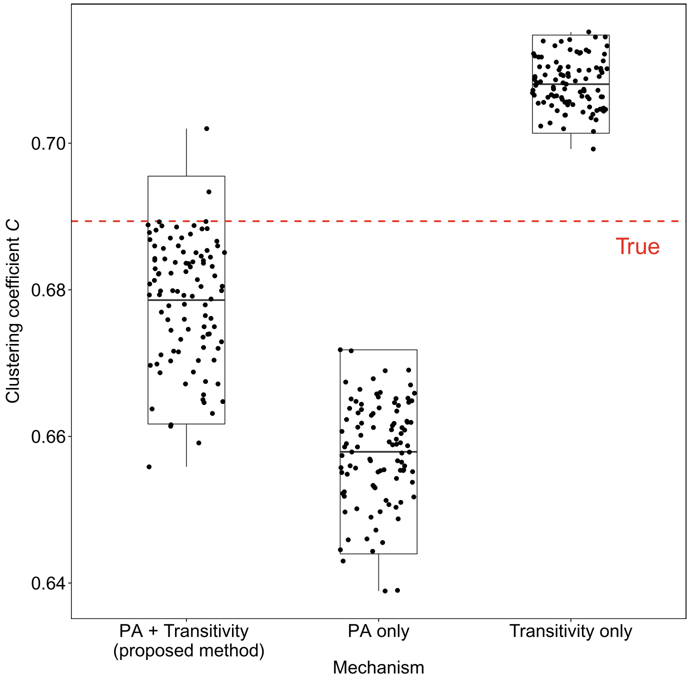


Fig. 1. Simulated clustering coefficients in three cases: the proposed method (jointly estimating the two functions), estimating the PA function in isolation, and estimating the transitivity function in isolation in the co-authorship network between scientists in the complex network field. The number of replications is 100. The plotted box in each case shows the population mean and its $\pm 2\sigma$ confidence interval. The red dashed line represents the value of the true clustering coefficient.

This implies that, if one wants to boost their number of collaborators, one should collaborate with well-connected scientists.

By quantifying the contribution of each mechanism in the growth process, we also found that while transitivity dominates PA in the high-energy physics network, preferential attachment triumphs transitivity in the remaining networks. Concluding remarks are given in Sect. 5.

2 Maximum Likelihood Estimation

In this section, we describe the maximum likelihood estimation method for the model described in Eq. (1). Let k_{max} be the maximum value of the degree of a node and b_{max} be the maximum value of the number of common neighbors between a pair of nodes. The variables we want to estimate are

$\boldsymbol{A} = [A_0, A_1, \ldots, A_{k_{max}-1}]$ and $\boldsymbol{B} = [B_0, B_1, \ldots, B_{b_{max}-1}]$. The log-likelihood function is:

$$l(\boldsymbol{A}, \boldsymbol{B}) = \sum_{t=1}^{T} \sum_{k_1=0}^{k_{max}-1} \sum_{k_2=k_1}^{k_{max}-1} \sum_{b=0}^{b_{max}-1} m_{k_1,k_2,b}(t) \log A_{k_1} A_{k_2} B_b - \sum_{t=1}^{T} m(t) \log \left(\sum_{k_1=0}^{k_{max}-1} \sum_{k_2=k_1}^{k_{max}-1} \sum_{b=0}^{b_{max}-1} n_{k_1,k_2,b}(t) A_{k_1} A_{k_2} B_b \right), \quad (2)$$

where $m(t)$ is the number of new edges at time-step t, $m_{k_1,k_2,b}(t)$ is the number of new edges at time-step t between node pairs (i, j) that satisfy $(k_i(t), k_j(t), b_{i,j}(t)) = (k_1, k_2, b)$, and $n_{k_1,k_2,b}(t)$ is the number of such pairs. By substituting $A_k = (k+1)^\alpha$ and $B_b = (b+1)^\beta$ in Eq. (2), we can obtain the log-likelihood function of the parametric model. In practice, logarithmic binning is applied to k and b for fast computation [13]. The log-likelihood function in Eq. (2) can be maximized by an MM algorithm [19]. Starting from some initial vectors $\boldsymbol{A}^0 = [A^0_0, A^0_1, \ldots, A^0_{k_{max}-1}]$ and $\boldsymbol{B}^0 = [B^0_0, B^0_1, \ldots, B^0_{b_{max}-1}]$, the algorithm updates $\boldsymbol{A}^i$ and $\boldsymbol{B}^i$ in parallel so that $l(\boldsymbol{A}^i,\ \boldsymbol{B}^i)$ is monotonically increasing in i.

3 Calculating the Contributions of PA and Transitivity

It is not straightforward to meaningfully quantify the contributions of PA and transitivity. According to Eq. (2), if we define $\hat{\boldsymbol{A}} = [\hat{A}_0, \hat{A}_1, \ldots, \hat{A}_{k_{max}-1}]$ and $\hat{\boldsymbol{B}} = [\hat{B}_0, \hat{B}_1, \ldots, \hat{B}_{b_{max}-1}]$ as the estimated PA and transitivity functions, then $(\hat{\boldsymbol{A}}, \hat{\boldsymbol{B}})$ and $(c\hat{\boldsymbol{A}}, d\hat{\boldsymbol{B}})$ give the same log-likelihood value for any positive constant factors c and d. This means that the magnitudes of $\hat{\boldsymbol{A}}$ and $\hat{\boldsymbol{B}}$ cannot be meaningfully compared with each other and summary statistics that bases solely on magnitudes such as the mean are also out of the question.

The contributions $v_{PA}(t)$ of PA and $v_{trans}(t)$ of transitivity at time-step t, however, can be meaningfully defined if we look at scale-invariant statistics, such as the variance of logarithmic values. From Eq. (1), we have:

$$\log P_{i,j}(t) = \log \hat{A}_{k_i(t)} \hat{A}_{k_j(t)} + \log \hat{B}_{b_{ij}(t)} + S(t),$$

for some constant $S(t)$ that is independent of i and j. We have the following three observations. Firstly, what matters is the variances in $\log P_{i,j}(t)$ across node pairs (i, j): the more uniform $\log P_{i,j}(t)$'s are, the closer the behavior at time-step t of the temporal network to that of an Erdös-Rényi random network where each pair (i, j) has the same probability of developing new edges regardless of $k_i(t)$, $k_j(t)$ and $b_{i,j}(t)$. Secondly, PA and transitivity mechanisms influence $\log P_{i,j}(t)$ through $\log \hat{A}_{k_i(t)} \hat{A}_{k_j(t)}$ and $\log \hat{B}_{b_{ij}(t)}$, respectively. Finally, since $\mathrm{Var}(\log(cX)) = \mathrm{Var}(\log(X) + \log c) = \mathrm{Var}(\log(X))$ for any positive constant factor c, the variances of $\log \hat{A}_{k_i(t)} \hat{A}_{k_j(t)}$ and $\log \hat{B}_{b_{i,j}(t)}$ are invariant to the scale of $(\hat{\boldsymbol{A}}, \hat{\boldsymbol{B}})$. From these observations, we define $v_{PA}(t)$ and $v_{trans}(t)$

to be the variances of $\log \hat{A}_{k_i(t)}\hat{A}_{k_j(t)}$ and $\log \hat{B}_{b_{i,j}(t)}$, respectively, where each variance is calculated from all node pairs that exist at time-step t. Note that $(\hat{\boldsymbol{A}}, \hat{\boldsymbol{B}})$ is estimated using all the time-steps as described in Sect. 2. A low $v_{PA}(t)$ or $v_{trans}(t)$ indicates that the contribution of the corresponding mechanism is weak at that time-step. For example, when $\hat{B}_b \approx 1$ for all b, i.e., there is almost no transitivity, then $\log \hat{B}_{b_{i,j}(t)}$ will take approximately the same value for all node pairs, and thus $v_{trans}(t)$ will be very close to zero. Although $\hat{A}_k$ and $\hat{B}_b$ do not change with t, $v_{PA}(t)$ and $v_{trans}(t)$ are inherently temporal for two reasons: (1) the set of all node pairs that exist at time-step t might be changing due to the births of new nodes, and (2) $\log \hat{A}_{k_i(t)}\hat{A}_{k_j(t)}$ and $\log \hat{B}_{b_{i,j}(t)}$ might be changing due to the changes of $k_i(t)$, $k_j(t)$ or $b_{i,j}(t)$.

4 Results

We analyzed three different scientific co-authorship networks: between authors in the complex network field (CMP) [20], authors of articles in the arXiv Hep-Th (high-energy theory) section (HEP) [21]), and authors of articles published in the SMJ (SMJ) [22]. While new collaborations and repeated collaborations are pooled together in HEP and SMJ, in CMP only new collaborations are considered. Table 1 shows the summary statistics for the three datasets.

Table 1. Summary statistics for three scientific co-authorship networks. $|V|$ and $|E|$ are the total numbers of nodes and edges in the final snapshot, respectively. T is the number of time-steps. $\Delta|V|$ and $\Delta|E|$ are the increments of nodes and edges after the initial snapshot, respectively. C is the clustering coefficient of the final snapshot.

Dataset	$\lvert V\rvert$	$\lvert E\rvert$	T	$\Delta\lvert V\rvert$	$\Delta\lvert E\rvert$	C
CMP	1498	2849	144	1377	2719	0.689
HEP	6798	290597	217	6796	290594	0.333
SMJ	2704	4131	354	2696	4123	0.378

We note that the number of edges per node in HEP is nearly 100 times greater than those in CMP and SMJ. As we shall see in Figs. 2a and b, this would cause the confidence interval at each estimated A_k and B_b in HEP to be very small comparing with those in CMP and SMJ.

Figures 2a shows the estimated PA functions in three networks. The estimated PA functions of CMP and SMJ are clearly increasing in average, which implies the existence of PA in these datasets. The estimated A_k of HEP is, however, only increasing up to $k_1 \approx 30$, from there it decreases up to $k_2 \approx 300$ and then once again increases. This might be caused by the fact that there are papers with hundreds of authors in high-energy physics and such papers are likely to be governed by some mechanism that is different than PA.

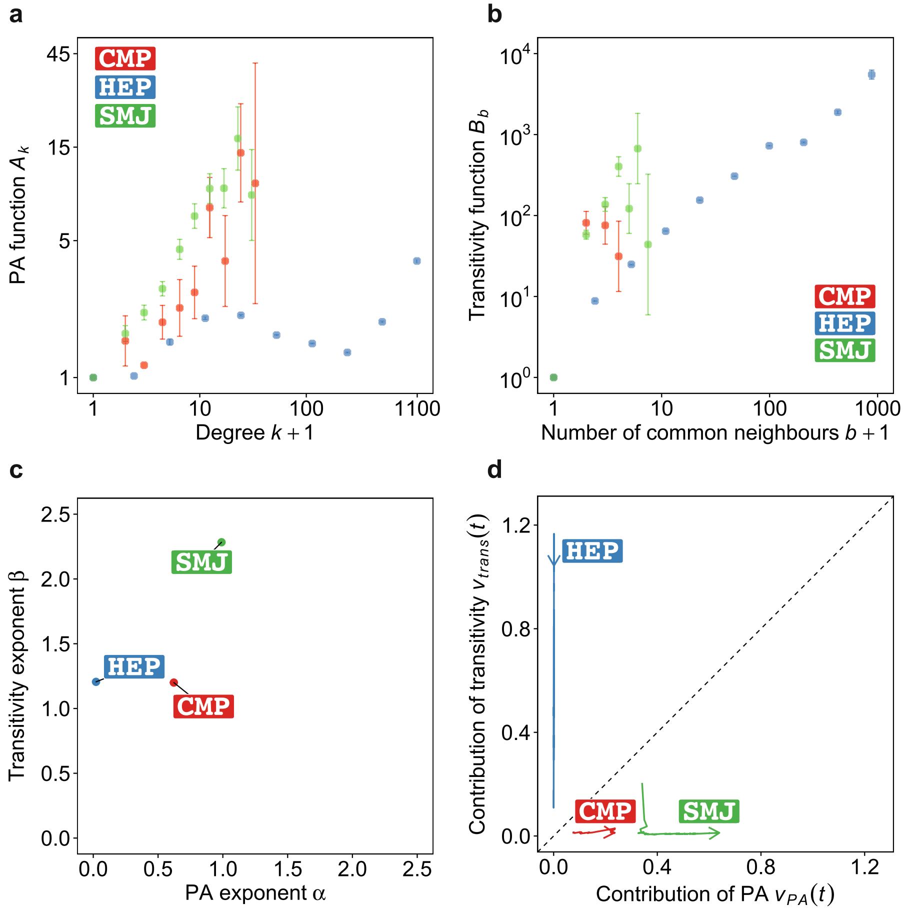


Fig. 2. Results in CMP, HEP, and SMJ networks. **a**: Estimated PA functions $\hat{A}_k$. The vertical bar at each point indicates the $\pm 2\sigma$ confidence interval. **b**: Estimated transitivity functions $\hat{B}_b$. The vertical bar at each point indicates the $\pm 2\sigma$ confidence interval. **c**: Estimated PA exponent α and transitivity exponent β. α and β are estimated by fitting the parametric forms $A_k = (k+1)^\alpha$ and $B_b = (b+1)^\beta$. **d**: The evolutions of $v_{PA}(t)$ and $v_{trans}(t)$. $v_{PA}(t)$ and v_{trans} are the contributions of PA and transitivity, respectively, at time-step t. The arrow in each dataset indicates the final time-step.

Figure 2b shows the estimated transitivity functions in three networks. In CMP, where only new collaborations are considered, the increase in B_1 from B_0 is nearly 100-fold. This is consistent with what was observed in Newman's analysis [14], which revealed a 50-fold increase in a co-authorship network from the Los Alamos E-print Archive. The implication of this observation is that one common co-author between a pair of scientists dramatically increase the chance that the two will collaborate in the future, thus one might want to co-author with well-connected scientists if one wants to boost their number of co-authors. When new collaborations and repeated collaborations are pooled together, B_1/B_0 is also very high: this quantity is about 10 and about 80 in HEP and SMJ, respectively, which also indicates the importance of one common collaborator in facilitating collaborations between a pair of authors.

When b is greater than 1, the confidence intervals in CMP and SMJ are too big to draw any conclusions beyond the observation that B_b in these two datasets fluctuate and show a non-increasing trend. The B_b in HEP, however, shows a clearly increasing trend when $b > 1$. This trend, together with Fig. 2d, shows the dominance of transitivity in HEP, which is consistent with the well-known fact that the field of high-energy physics enjoys many long-lasting collaborations [23, 24].

Although the non-parametric A_k and B_b in these datasets seem to not follow any simple parametric forms, it is still beneficial to consider the power-law forms $A_k = (k+1)^\alpha$ and $B_b = (b+1)^\beta$ and then estimate the PA exponent α and transitivity exponent β. These exponents reveal the general trend of the two mechanisms and might be important in predicting the future of the network. Figure 2c shows the estimated α and β in three networks. The estimated values of α are founded to be from 0 to 1, which are consistent with previous works [13, 14]. This range of α falls in the linear and sub-linear PA regions, which have been shown to lead to stable degree distributions when there is PA alone [12]. Extrapolating this result to our case, it is likely that some stable degree distribution would arise in the long run, too. This is important since it means that the low-degree authors would still have a chance to develop new collaborations, as opposite to the "winner-take-all" situation when the degree distribution is not stable. The estimated values of β are all greater than 1. There is, however, no study on the effects of such values of β on the asymptotic behavior of the network.

Finally, Fig. 2d shows $v_{PA}(t)$ and $v_{trans}(t)$ of three datasets at each time-step. In HEP, $v_{trans}(t)$ is greater than $v_{PA}(t)$ for all time-steps t, which suggests that transitivity completely dominates PA in this network. In CMP and SMJ networks, the situation is reversed: $v_{PA}(t)$ is greater than $v_{trans}(t)$ for all t, which implies that PA triumphs transitivity in CMP and SMJ.

5 Concluding Remarks

We proposed a non-parametric joint estimation method for the PA and transitivity mechanisms. Applying our method to three scientific co-authorship networks, we found the main driving force behind the evolution of each network is different: it is transitivity in Hep-Th network, and is PA in the remaining two. Furthermore, we found that B_1 increases at least five fold from B_0, which indicates the importance of one common collaborators in facilitating new collaborations between two authors. Among many possible directions for future research, two particularly stand out. The first one is to incorporate the fitness mechanism [25] to express node heterogeneity which is ignored in the PA and transitivity mechanisms. The second promising direction is to explore the evolutions of the PA and transitivity mechanisms themselves by investigating how the functions A_k and B_b change with time.

References

1. Nowak, M.A.: Five rules for the evolution of cooperation. Science **314**(5805), 1560–1563 (2006). http://science.sciencemag.org/content/314/5805/1560
2. Dugatkin, L.A.: Cooperation Among Animals: An Evolutionary Perspective. Oxford University Press, Oxford (1997)
3. Watson, A.: Diplomacy. Routledge, London (1984)
4. Hamel, G., Doz, Y.L., Prahalad, C.K.: Collaborate with your competitors-and win. Harv. Bus. Rev. **67**(1), 133–139 (1989). https://hbr.org/1989/01/collaborate-with-your-competitors-and-win
5. Johnson, D.W., Johnson, R.T., Smith, K.A.: Active Learning: Cooperation in the College Classroom. Interaction Book Company, Edina (1991)
6. Larivière, V., Gingras, Y., Sugimoto, C.R., Tsou, A.: Team size matters: collaboration and scientific impact since 1900. J. Assoc. Inf. Sci. Technol. **66**(7), 1323–1332. https://onlinelibrary.wiley.com/doi/abs/10.1002/asi.23266
7. Bornmann, L.: Is collaboration among scientists related to the citation impact of papers because their quality increases with collaboration? An analysis based on data from F1000Prime and normalized citation scores. J. Assoc. Inf. Sci. Technol. **68**(4), 1036–1047 (2017). https://doi.org/10.1002/asi.23728
8. Tahai, A., Meyer, M.J.: A revealed preference study of management journals' direct influences. Strateg. Manag. J. **20**(3), 279–296 (1999). https://onlinelibrary.wiley.com/doi/abs/10.1002/%28SICI%291097-0266%28199903%2920%3A3%3C279%3A%3AAID-SMJ33%3E3.0.CO%3B2-2
9. Holme, P., Kim, B.J.: Growing scale-free networks with tunable clustering. Phys. Rev. E **65**, 026107 (2002)
10. Newman, M.E.J.: Coauthorship networks and patterns of scientific collaboration. In: Proceedings of the National Academy of Sciences, vol. 101(suppl 1), pp. 5200–5205 (2004). http://www.pnas.org/content/101/suppl_1/5200
11. Newman, M.E.J.: Scientific collaboration networks. I. network construction and fundamental results. Phys. Rev. E **64**, 016131 (2001). https://link.aps.org/doi/10.1103/PhysRevE.64.016131
12. Krapivsky, P., Rodgers, G., Redner, S.: Degree distributions of growing networks. Phys. Rev. Lett. **86**(23), 5401–5404 (2001)
13. Pham, T., Sheridan, P., Shimodaira, H.: PAFit: a statistical method for measuring preferential attachment in temporal complex networks. PLoS ONE **10**(9), e0137796 (2015)
14. Newman, M.: Clustering and preferential attachment in growing networks. Phys. Rev. E **64**(2), 025102 (2001)
15. Jeong, H., Néda, Z., Barabási, A.: Measuring preferential attachment in evolving networks. Europhys. Lett. **61**(61), 567–572 (2003)
16. Ripley, R., Boitmanis, K., Snijders, T.A.: RSiena: Siena - Simulation Investigation for Empirical Network Analysis, R package version 1.1-232 (2013). https://CRAN.R-project.org/package=RSiena
17. Krivitsky, P.N., Handcock, M.S.: tergm: Fit, Simulate and Diagnose Models for Network Evolution Based on Exponential-Family Random Graph Models. The Statnet Project. R package version 3.4.0 (2016). http://www.statnet.org, http://CRAN.R-project.org/package=tergm
18. Kong, J., Sarshar, N., Roychowdhury, V.: Experience versus talent shapes the structure of the web. Proc. Nat. Acad. Sci. U.S.A. **37**, 105 (2008)

19. Hunter, D., Lange, K.: Quantile regression via an MM algorithm. J. Comput. Graph. Stat. **9**, 60–77 (2000)
20. Pham, T., Sheridan, P., Shimodaira, H.: PAFit: an R Package for the Non-parametric Estimation of Preferential Attachment and Node Fitness in Temporal Complex Networks. ArXiv e-prints, April 2017
21. KONECT: arxiv hep-th network dataset. http://konect.uni-koblenz.de/networks/ca-cit-HepTh. Accessed 03 May 2018
22. Ronda-Pupo, G.A., Pham, T.: The evolutions of the rich get richer and the fit get richer phenomena in scholarly networks: the case of the strategic management journal. Scientometrics, May 2018. https://doi.org/10.1007/s11192-018-2761-3
23. Zimmermann, F.: High-energy physics strategies and futurelarge-scale projects. Nucl. Instr. Meth. Phys. Res. Sect. B: Beam Interact. Mater. Atoms **355**, 4–10 (2015). http://www.sciencedirect.com/science/article/pii/S0168583X1500350X. Proceedings of the 6th International Conference Channeling 2014: Charged & Neutral Particles Channeling Phenomena, 5–10 October 2014, Capri, Italy
24. Birnholtz, J.P.: What does it mean to be an author? The intersection of credit, contribution, and collaboration in science. J. Am. Soc. Inf. Sci. Technol. **57**(13), 1758–1770. https://onlinelibrary.wiley.com/doi/abs/10.1002/asi.20380
25. Pham, T., Sheridan, P., Shimodaira, H.: Joint estimation of preferential attachment and node fitness in growing complex networks. Sci. Rep. **6** (2016). https://doi.org/10.1038/srep32558

The Impact of New Mobility Modes on a City: A Generic Approach Using ABM

Arnaud Grignard[1](✉), Luis Alonso[1], Patrick Taillandier[2], Benoit Gaudou[3], Tri Nguyen-Huu[4,6], Wolfgang Gruel[5], and Kent Larson[1]

[1] MIT Media Lab, City Science, Cambridge, USA
agrignard@gmail.com
[2] MIAT, University of Toulouse, INRA, Toulouse, France
[3] University Toulouse 1 Capitole, UMR 5505 IRIT, CNRS, Toulouse, France
[4] UMMISCO, IRD, Bondy, France
[5] Institute for Mobility and Digital Innovation, Stuttgart Media University, Stuttgart, Germany
[6] IXXI, ENS Lyon, Lyon, France

Abstract. Mobility is a key issue for city planners. Being able to evaluate the impact of its evolution is complex and involves many factors including new technologies like electric cars, autonomous vehicles and also new social habits like vehicle sharing. We need a better understanding of different scenarios to improve the quality of long-term decisions. Computer simulations can be a tool to better understand this evolution, to discuss different solutions and to communicate the implications of different decisions. In this paper, we propose a new generic model that creates an artificial micro-world which allows the modeler to create and modify new mobility scenarios in a quick and easy way. This not only helps to better understand the impact of new mobility modes on a city, but also fosters a better-informed discussion of different futures. Our model is based on the agent-based paradigm using the GAMA Platform. It takes into account different mobility modes, people profiles, congestion and traffic patterns. In this paper, we review an application of the model of the city of Cambridge.

Keywords: Agent-based modeling · Mobility · City Science

1 Introduction

In recent years, we have seen significant changes in the way we travel. These changes are partially driven by new technologies like electric vehicles, but also by new usage concepts such as car-sharing or ride-hailing [8]. It is anticipated that these changes will accelerate in the coming years - with wide-spread adoption of shared, autonomous and electric vehicles. For cities, it is crucial to proactively manage this development. However, this task is challenging because of the speed

A. J. Morales et al. (Eds.): ICCS 2018, SPCOM, pp. 272–280, 2018.
https://doi.org/10.1007/978-3-319-96661-8_29

of the development and because the mobility system has many dependencies e.g., to the cities' infrastructure or its economic and social systems. Simulations can be powerful tools to test and to communicate the impact of decisions in the field of mobility. Numerous simulation tools exist to investigate current traffic questions, but few are able to examine the impact of future mobility systems. Most of the existing tools focus on specific problems and are limited to existing mobility modes.

Using meaningful simulation models might help to better understand and evaluate the effects of new services on different stakeholders in cities. Service operators might use these simulations to support the development-process of their services, cities can use them to experiment with regulatory variables like subsidies, taxes, or congestion charges, while citizens can investigate the effects on traffic-flows in their neighborhood. New mobility options are currently being developed and neither their characteristics nor the behavioral changes they create can be measured or predicted in an exact way. That is why methods to investigate the impact of these options need to deal with a high-level of uncertainty. A method to reduce uncertainty is to integrate a multitude of stakeholders into the discussion. Modeling systems are needed to quickly create, change and test different scenarios.

We propose an agent-based model (ABM) that allows the modeler to easily integrate new transportation modes. The main goal of this model is to show the impact of different mobility modes on traffic flow and congestion. We aim to address other aspects like housing or energy flows in the future. The model, implemented with the GAMA modeling platform [6], is developed as an open source project that also allows the modeler to easily develop and change mobility scenarios. We describe a framework which can be adapted easily for different contexts and we describe its application in a case study. The focus of our work is on the quick development of meaningful scenarios and the support of discussions between stakeholders rather than creating exact modeling results. The paper is structured as follows: Sect. 2 presents the context of this work. Section 3 is dedicated to the presentation of the model using the ODD protocol [7]. Section 4 presents a case study and its output. Section 5 concludes and presents perspectives.

2 Limits of the Existing Tools for Traffic Simulation

To investigate the complex relationships of mobility, a multitude of tools are used. Among them System Dynamics is used to investigate complex problems, to model the relationship between transportation systems [13], and to explore the impact of new systems like autonomous vehicles [9]. Questions of current and future mobility systems are also examined with the help of mathematical methods [2,12]. However Agent Based Modelling remains one the most popular approaches and is the base to many frameworks for traffic simulation. Among them MATSim [3] is an open-source platform with different features dedicated to traffic simulation which can be defined through the use of new modules in JAVA.

SUMO [11] proposes many advanced features that can be developed and enriched using C++. However, for modelers without high-level programming skills, adapting these platforms to specific application contexts, such as the integration of new transport modes, is very difficult and out of reach. As a consequence, many models are still developed from scratch (e.g. [10,15])) which is time consuming and limits the re-usability of the models.

In order to face this difficulty, some authors [4,5] used a generic modeling platform to develop their models. The interest is to facilitate the comparison of models and their re-usability. One of the existing generic platforms that is particularly adapted to develop transportation models is the GAMA platform [6]. GAMA provides modelers with a complete modeling language and an integrated development environment that permits the user to build models in a rapid and easy manner. The GAMA platform provides different features that can be used by modelers to develop transportation models. In particular, it is possible to simply load GIS data, define graphs from polyline geometries, compute the shortest paths and move agents on a network. Additionally, some movement behaviors are already pre-coded and can be easily reused. Many transportation models were produced in the last years with the GAMA platform [14]. However, these models have not focused on the impact of future mobility options and they have not provided a generic framework to investigate potential impacts of these options. We use the GAMA platform to create a framework that can be adapted easily to investigate how the discussion about possible solutions is changed by models that are adaptable during that discussion. Our focus is to stimulate a well-founded discussion around different scenarios and not to model the situation as exact as possible. The discussion is supposed to generate promising scenarios that are meant to be investigated more in detail.

3 Model Description

Our model aims to simulate the mobility-behavior of city dwellers and create meaningful output for the discussion around new mobility services. It offers different configurations and focuses on three main categories of urban travel: non motorized transport, public transit and private motorized transport. This allows us to investigate the impact of new mobility modes and to observe behavioral changes over time considering different plans. It also reveals how urban forms and transportation infrastructure impact travel patterns.

3.1 Entities, State Variables, and Scales

We list here the main variables used in the model. The environmental variables are the following:

– **Transport Network**: A graph on which agents can travel during the simulation. Network edges have the following attributes: *type*: the type of infrastructure (road, rail), *max speed*: the maximum authorized speed on the network

segment, *capacity*: the capacity in terms of agents (people, cars), proportional to the length of the network segment, *current concentration*: the number of people on the network segment.
- **Building**: each building is represented by a polygon and has three attributes: *usage* (Residential or Office), *scale* (Small (S), Medium (M), or Large (L)), *category* (Restaurant, Nightlife, Gathering Places, Cultural, Shopping, High School, University, Office, Residential, or Park).
- **Mobility Hub**: a place where people can enter/exit a mobility mode, such as a bus stop or train station. It contains one single attribute: *waiting_people*, the list of people waiting at this mobility hub.

The agents species used in the models are:

- **People**: People have several activities during the day and move around the city accordingly. People agents' behavior is to determine the location of the next activity, choose a mobility mode and move towards that location. Their attributes include: *profile*: the profile of the agent (e.g. High School Student, Executives, Retirees), *destination*: the final destination that the agent wants to reach, *living_place*: its living place, *objectives*: the schedule of objectives during the day, *current_objective*: its current objective, *mobility_mode*: the mobility mode currently used, *possible_mobility_modes*: the list of the mobility modes that are available to that agent.
- **Transit Vehicle**: Public transit vehicles follow a route and can embark/disembark people. This species has 3 attributes: *stops*: the list of stops, *stop_passengers*: the list of people waiting at each stop, *my_target*: the current target of the vehicle (iteration of the list of stops).

3.2 Process Overview and Scheduling

The dynamics of the model is based on 3 consecutive steps: (1) management of traffic jams by the network agents, (2) behavior of the transit agents and (3) behavior of the people agents. Each step of the simulation represents 1 min. The management of the traffic jam is based on the following rule: the more people agents move on a network segment, the lower their speed. The transit vehicle agents have a simple behavior: according to their schedules, they move from hub to hub and pick up the people agents that are waiting at each station. The behavior of the people agents is defined by their list of objectives which defines their schedule. Figure 1 gives an example: an agent wakes up, (1) goes to work and stays there for a while, (2) goes to lunch, (3) gets back to work, (4) goes to dinner and finally (5) travels back home.

Whenever an agent changes its activity, it may need to move to a new location, determine a transportation mode and compute the path to its destination. Factors like travel time or price can influence the agent's mode choice. The agents may weigh those factors differently, depending on their profile and the type of activity.

The computation of the mode of transportation is performed each time an agent decides to move. The algorithm consists of three steps: (i) Determination of the possible modes of transportation. This step relies on simple considerations, such as whether or not agent owns a car. (ii) Computation of factors related to each of possible mode. Currently, we consider the following factors: travel-time, price, and difficulty. Difficulty represents amount of effort that is needed: walking to a mobility hub and waiting may be seen as more difficult and time consuming than using one's own car. (iii) Choice of the mode of transportation. Each agent will determine its mode of transportation, based on its own characteristics (profile), the purpose of the trip and the indicators computed at step 2.

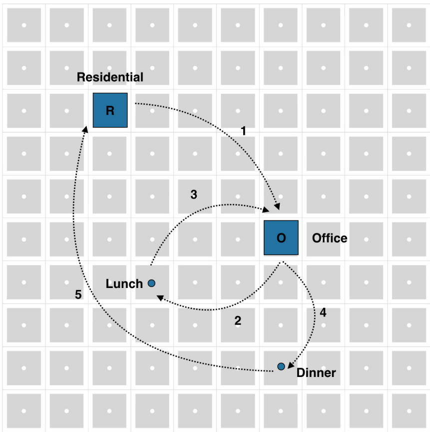


Fig. 1. Model process

Step (iii) consists in attributing a score to each transportation mode, according to a Travel Route Initialization formula based on a linear combination of the considered factor, with weights depending on the context of the travel (social category of the agent, purpose of the travel). For example, a student is more likely to use a bike than a retired person, hence the student will have a greater weight for using a bike. This Travel Route Initialization formula entirely relies on a table of weights that need to be determined for each case study. It provides the following advantages: firstly, it relies on simple mathematical considerations that are understandable by end users and stakeholders. Secondly, it provides flexibility, as it is possible to add, remove or modify the factors considered. Finally, it is easy to compute the table of weights from the data, either from individual surveys or raw mobility data. In the later case, the weight table can be determined by a multinomial logistic regression based with chosen predictors. All of this makes this approach a general framework that can easily modified and adapted to suit the requirements for different case studies.

We present here a small example with two profiles: college student and mid-carrier workers. The values in Table 1a represent how important each category is. A positive (resp. negative) value means a positive (resp. negative) influence. Here the cost of travel has a very negative influence for the college student, while

Table 1. Example for two profiles and an imaginary trip

Profile	Price	Time	Difficulty
College student	-0.9	-0.2	-0.65
Mid-carrier worker	0	-1	-1

(a) Weights

Mode	Bike	Car	Bus
Price	0	8	1.5
Time	40	10	20
Difficulty	0.1	0.2	0.5

(b) Example values

it does not matter at all for the mid-carrier worker. Table 1b shows values for an imaginary trip.

From those tables, we compute a score with the above formula, shown in 2. Each agent will chose the mode corresponding to the best score, i.e. bus for the student and car for the mid-carrier worker.

Table 2. Computed scores

Profile	Bike	Car	Bus
College student	−8.065	−9.33	−5.65
Mid-carrier worker	−40.1	−10.2	−20.5

3.3 Initialization and Input Data

To initialize the model, the following steps have to be taken into account: (1) the building and mobility network agents are created from GIS files (shapefiles); (2) the different mobility modes are generated from a csv file that provides information about price, time traveled (speed, wait-time), and difficulty of use; (3) the people agents are created by reading two csv files: one that gives information about the agents' distribution and the probabilities to have the different mobility modes (e.g. bike or car). The second one contains information about the agents' typical schedules (activities for each hour); (4) the weights per profile and per activity are imported from a csv file; (5) the people agents are created and their location is determined based on data from the profile files.

4 Case Study: City of Cambridge - Kendall Square

The model described early in this paper has been applied in a larger project called CityScope Volpe developed by the City Science group at the MIT Media Lab [1] as shown in Figs. 2 and 3. MIT was recently chosen as the partner to redevelop the 14-acre site of Volpe, which formerly housed the U.S Department of Transportation. CityScope is a platform for shared, interactive computation for urban planning. It uses LEGO bricks as a tangible interface to the computational models and aims to interactively engage stakeholders in the city-planning-process.

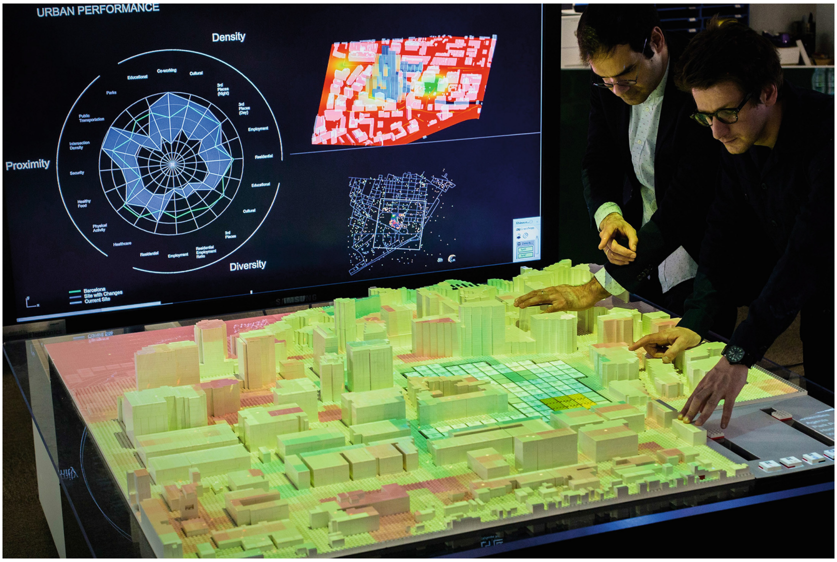


Fig. 2. CityScope: a data-driven interactive simulation tool for urban design.

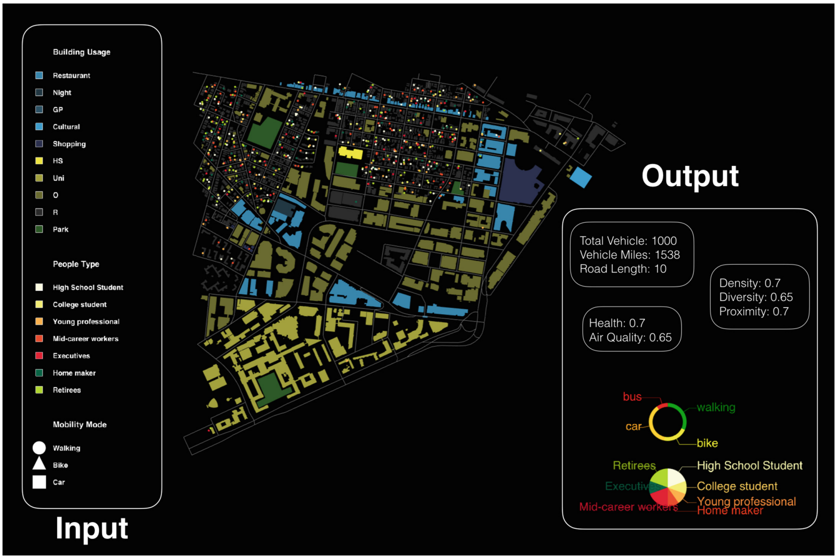


Fig. 3. Model overview.

Walkability, energy or economic activity models are used on the CityScope platform to solve spatial design and urban planning challenges. The mobility model focuses on three main aspects: total number of vehicles, vehicle miles and hours traveled, road length and congestion. The CityScope platform was used in several workshops in order to discuss different suggestions for the site. We applied the process that has been described before to initialize the model. For the building and mobility network (step 1), we used the proposals from the design stage of the project. For the mobility modes (step 2), we adapted information that is available from the Massachusetts Bay Transportation Authority. The people agents (step 3) could be configured by the stakeholders during the workshops. This allows for different ideas about zoning and density to be discussed. For the weights per profile, we adapted census-data for the Boston area (step 4). In our workshops, the stakeholders could change different parameters of our model by using the interactive LEGO interface. A change of the parameters resulted in a re-computation of the model.

5 Conclusion and Discussion

In the paper, we described a generic framework to investigate the impact of mobility modes on different aspects of daily life. It was designed to be easily adaptable to all cities and to easily integrate new mobility modes to investigate their possible impacts. Our case study showed that without programming-skills it is possible to create an ABM model that can be used to create meaningful output for the discussion around the design of mobility solutions. As our models come with many simplifications and focus on an unknown future, many uncertainties remain. However, the workshops revealed that the models contribute to a better understanding of the mobility system, even it is unclear how well specified the models need to be in order to take the optimal decision. It can be assumed that this is different for every project, and that it makes sense to refine the models during the process and to conduct in-depth investigations with different methods for the most promising alternatives. Although the creation of the model was possible without programming-skills, creating a mobility model is still not a trivial task. In order to integrate more people in the discussion about livable cities and to leverage the wisdom of crowds, the process needs to be further simplified. The stakeholders that took part in our workshops accepted our models. Despite the limitations of the model, workshop participants valued the impression of future scenarios, complex relationships became obvious and changing input parameters had direct input on the scenarios. They assume that this could lead to better decision making, than decisions that are either driven by gut-feelings or by experts who use tools that are hard to understand.

References

1. Alonso, L., Zhang, Y., Grignard, A., Noyman, A., Sakai, Y., ElKatsha, M., Doorley, R., Larson, K.: Cityscope: a data-driven interactive simulation tool for urban design. Use case volpe. In: ICCS 2018 (2018, to be Published)
2. Alonso-Mora, J., Samaranayake, S., Wallar, A., Frazzoli, E., Rus, D.: On-demand high-capacity ride-sharing via dynamic trip-vehicle assignment. Proc. Natl. Acad. Sci. **114**, 462–467 (2017)

3. Balmer, M., Rieser, M., Meister, K., Charypar, D., Lefebvre, N., Nagel, K., Axhausen, K.: MATSim-T: architecture and simulation times. In: Multi-Agent Systems for Traffic and Transportation Engineering, pp. 57–78 (2009)
4. Czura, G., Taillandier, P., Tranouez, P., Daudé, É.: Mosaiic: city-level agent-based traffic simulation adapted to emergency situations. In: Proceedings of the International Conference on Social Modeling and Simulation, plus Econophysics Colloquium 2014, pp. 265–274. Springer (2015)
5. Fosset, P., Banos, A., Beck, E., Chardonnel, S., Lang, C., Marilleau, N., Piombini, A., Leysens, T., Conesa, A., Andre-Poyaud, I., et al.: Exploring intra-urban accessibility and impacts of pollution policies with an agent-based simulation platform: GaMiroD. Systems **4**(1), 5 (2016)
6. Grignard, A., Taillandier, P., Gaudou, B., Vo, D.A., Huynh, N.Q., Drogoul, A.: GAMA 1.6: advancing the art of complex agent-based modeling and simulation. In: International Conference on Principles and Practice of Multi-Agent Systems, pp. 117–131. Springer (2013)
7. Grimm, V., Berger, U., DeAngelis, D.L., Polhill, J.G., Giske, J., Railsback, S.F.: The ODD protocol: a review and first update. Ecol. Model. **221**(23), 2760–2768 (2010)
8. Gruel, W., Piller, F.: New vision for personal transportation. MIT Sloan Manage. Rev. **57**(2), 20–24 (2016)
9. Gruel, W., Stanford, J.M.: Assessing the long-term effects of autonomous vehicles: a speculative approach. Transp. Res. Procedia **13**, 18–29 (2016). http://www.sciencedirect.com/science/article/pii/S2352146516300035, towards future innovative transport: visions, trends and methods 43rd European Transport Conference Selected Proceedings
10. Horn, M.: Multi-modal and demand-responsive passenger transport systems: a modelling framework with embedded control systems. Transp. Res. Part A Policy Pract. **36**, 167–188 (2002)
11. Krajzewicz, D., Erdmann, J., Behrisch, M., Bieker, L.: Recent development and applications of SUMO - Simulation of Urban MObility. Int. J. Adv. Syst. Meas. **5**(3&4), 128–138 (2012)
12. Pavone, M., Smith, S.L., Frazzoli, E., Rus, D.: Robotic load balancing for mobility-on-demand systems. Int. J. Robot. Res. **31**(7), 839–854 (2012)
13. Shepherd, S.: A review of system dynamics models applied in transportation. Transp. B Transp. Dyn. **2**(2), 83–105 (2014)
14. Taillandier, P.: Traffic simulation with the gama platform. In: International Workshop on Agents in Traffic and Transportation, pp. 8–p (2014)
15. Tranouez, P., Daudé, E., Langlois, P.: A multiagent urban traffic simulation. J. Nonlinear Syst. Appl. **3**(2), 98–106 (2012)

Quantifying Humans' Priors Over Graphical Representations of Tasks

Gecia Bravo Hermsdorff(✉), Talmo Pereira, and Yael Niv

Princeton Neuroscience Institute, Princeton University, Princeton, NJ 08544, USA
geciah@princeton.edu

Abstract. Some new tasks are trivial to learn while others are almost impossible; what determines how easy it is to learn an arbitrary task? Similar to how our prior beliefs about new visual scenes colors our perception of new stimuli, our priors about the structure of new tasks shapes our learning and generalization abilities [2]. While quantifying visual priors has led to major insights on how our visual system works [5,10,11], quantifying priors over tasks remains a formidable goal, as it is not even clear how to define a task [4]. Here, we focus on tasks that have a natural mapping to graphs. We develop a method to quantify humans' priors over these "task graphs", combining new modeling approaches with Markov chain Monte Carlo with people, MCMCP (a process whereby an agent learns from data generated by another agent, recursively [9]). We show that our method recovers priors more accurately than a standard MCMC sampling approach. Additionally, we propose a novel low-dimensional "smooth" (In the sense that graphs that differ by fewer edges are given similar probabilities.) parametrization of probability distributions over graphs that allows for more accurate recovery of the prior and better generalization. We have also created an online experiment platform that gamifies our MCMCP algorithm and allows subjects to interactively draw the task graphs. We use this platform to collect human data on several navigation and social interactions tasks. We show that priors over these tasks have non-trivial structure, deviating significantly from null models that are insensitive to the graphical information. The priors also notably differ between the navigation and social domains, showing fewer differences between cover stories within the same domain. Finally, we extend our framework to the more general case of quantifying priors over exchangeable random structures.

Keywords: Markov chain Monte Carlo with People (MCMCP) · Representational learning · Structural priors · Task graphs · Human cognition

1 Introduction

1.1 Our Brain Must Utilize Efficient Priors

Our lives are punctuated by a multitude of seemingly disparate new tasks (e.g., navigating in a new place, interacting with new people, writing a new abstract)

A. J. Morales et al. (Eds.): ICCS 2018, SPCOM, pp. 281–290, 2018.
https://doi.org/10.1007/978-3-319-96661-8_30

that we are able to perform with relative ease. Still, if we consider all tasks we could possibly be faced with, we would not be good (at least initially) at most of them (e.g., playing Go). This simple observation leads to a fundamental, yet unanswered, question in cognitive neuroscience: are there essential structural properties that unite the tasks that our brains are "naturally" good at solving, and if so, what are they?

Understanding our brain's representation of a new task (i.e., our prior about the task's structure) is key to answering this question. Indeed, the prior used in a given task sharply constrains how fast and efficiently (if at all) this task can be solved [2]. In particular, the curse of dimensionality [3] suggests that task representations should be compact, filtering out redundancies. However, there is no free lunch; reduced representations also constrain the set of tasks an agent can efficiently solve. Thus, these reduced representations should manifest as priors that leverage on the relevant structure of naturalistic tasks, i.e., tasks that the organism encounters in everyday life and have been relevant over evolutionary time-scales [4].

1.2 Main Contributions

Quantifying priors over tasks is a formidable goal [2,4], if not only for the reason that "what is a task?" is a relatively open-ended question. Here we restrict our attention to tasks that have a natural mapping to graphs as this allows us to quantify their structure using graph theoretical tools. Specifically, our experiments focus on two prominent domains of naturalistic tasks: navigation and social interaction, with nodes and edges representing, for example, regions and borders, or people and relationships.

On the theoretical side, we develop a method to quantify humans' priors over these "task graphs", which combines new modeling approaches with Markov chain Monte Carlo with people (MCMCP) [6,9] – a process whereby an agent learns from data generated by another agent, recursively. Our simulations demonstrate that our method recovers these priors more accurately than a standard MCMC sampling approach. This result is particularly relevant for the "resource constrained" regime, where data are limited and costly to acquire, such as in experiments with human subjects. Moreover, we propose a novel low-dimensional "smooth" parametrization of probability distributions over (non-isomorphic) graphs on the same vertex set. We show that, in the limited data regime, it allows for more accurate recovery of the prior (*in silico* data), and better generalization (in human data). Finally, we extend our framework to the more general case of quantifying priors over exchangeable random structures [14].

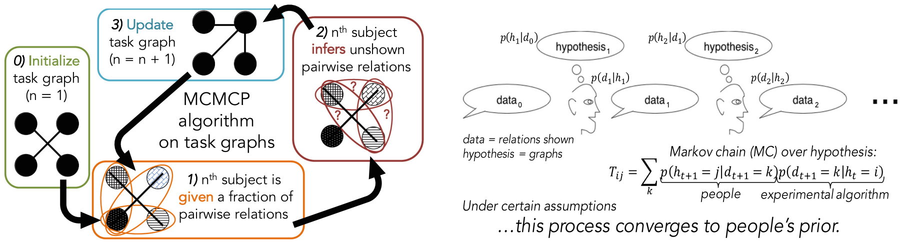
On the experimental side, we have created an online experimental platform[1] with a game-like interface that instantiates our MCMCP algorithm, and allows

[1] Links for some of our experiments:
http://psiturk-geciah.princeton.edu:9001/ (navigation in nature parks),
http://psiturk-geciah.princeton.edu:9003/ (navigation in cities),
http://psiturk-geciah.princeton.edu:9000/ (friendships in workplaces),
http://psiturk-geciah.princeton.edu:9002/ (friendships in school classes).

subjects to interactively draw the task graphs. We use this platform to collect human data on several navigation and social interactions tasks. We show that priors over these tasks have non-trivial structure, deviating significantly from null models that are insensitive to the graphical information. Moreover, the priors are notably different between the navigation and social domains, while exhibiting fewer differences between different tasks in the same domain (city and nature park (navigation); coworkers and students (social)).

2 Markov Chain Monte Carlo with People (MCMCP) over (task) Graphs

Figure 1a illustrates our MCMCP algorithm for generating experiments. For example, in one of our experiments, the subject is told that they are visiting a new city, and informed whether certain pairs of neighborhoods share a border or not (*step 1* in Fig. 1a). They then are asked to guess if the other pairs of neighborhoods share a border or not (*step 2*) by drawing a map using our graph drawing interface. Additionally, to incentivize subjects to give their true prior, they are told that there is an underlying truth (e.g., an actual city), and that they win extra money by correctly guessing the relations obscured.



(a) (b)

Fig. 1. *Performing Markov chain Monte Carlo with people (MCMCP) in a given task allows for sampling from their prior over it.* (a) *Schema of our algorithm for generating experiments: (0)* create a task graph for the first subject; *(1)* the subject is given partial information about this graph (a fraction of the pairwise relations chosen at random, here, 3 out of 6); *(2)* they are asked to infer the unshown portions (the remaining pairwise relations); *(3)* construct a new task graph from these responses for the next subject; *(4)* repeat steps *1–3*. (b) *The algorithm interpreted as MCMC.*

This back-and-forth between data seen by the subjects (relations shown, or "partial graphs") and the resulting hypothesis they infer (completed "task graphs") can be marginalized over the partial graphs to create a Markov chain (MC) over the space of task graphs (Fig. 1b). Assuming that subjects are Markovian, and share the same fixed decision rule, this MC is time-homogeneous. If, in addition, we assume that subjects are "Bayesian", computing Bayes rule using the correct likelihood function[2] and a shared prior, and respond by sampling from their posterior, this MC has as its stationary distribution the subjects' prior over the relevant task graphs.[3] Precisely, this "MCMCP Bayesian model" gives a transition matrix $\underline{\underline{\mathbf{T}}}$ over the relevant non-isomorphic task graphs, with entries:

$$t_{ij} = p(g_j|g_i) = \sum\nolimits_k p(g_j|d_k)p(d_k|g_i) \tag{1}$$

where $p(d_k|g_i)$ is the probability of seeing partial graph d_k by randomly obscuring r relations of the graph g_i (with $r :=$ total number of relations minus number of relations shown), and $p(g_j|d_k)$ is given by Bayes rule using a fixed prior.

3 Results

3.1 Resource Constrained MCMCP

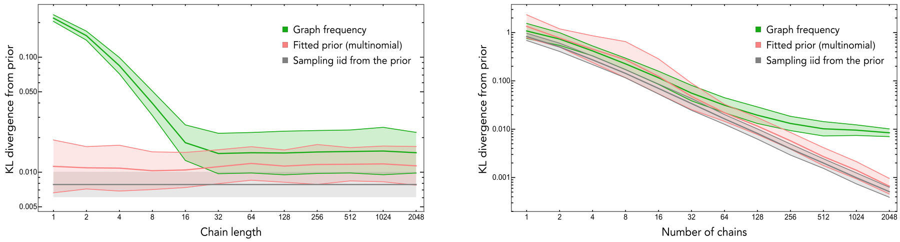
In standard MCMC, one uses samples generated by the algorithm to reconstruct the target (stationary) distribution. This is inefficient in the following sense: to obtain i.i.d. samples, only a small fraction of the iterations are used as samples, as one must discard the initial samples until (hopefully[4]) the chain has converged to its stationary distribution (the so called "burn-in" period). In addition, one might only collect samples every $\mathcal{O}(\tau_c)$ iterations (where $\mathcal{O}(\tau_c)$ is the autocorrelation time) to mitigate correlations. While these issues are generally not a concern when samples are efficiently generated via a computer, in MCMCP the primary bottleneck is due to the use of human subjects.

Fortunately, in our case, we can use data more efficiently by leveraging the additional structure provided by the Bayesian assumption. Specifically, we propose to recover subjects' prior by fitting their choices to our MCMCP Bayesian model (as opposed to using the observed graph frequency as a proxy of the prior as is done in classic MCMC). The unknowns are the probabilities that the prior gives to each of the non-isomorphic graphs on the relevant vertex set. As illustrated in Fig. 2, our fitting method recovers the prior more precisely than a standard MCMC sampling method (especially in the case of constrained chain length). Moreover, aside from using data more efficiently, our approach has additional advantages: it does not have the problem of "guessing" the mixing time (which can vary substantially depending on the prior, number of nodes, and number of relations obscured); and it also allows for experiments to be run in parallel.

[2] I.e., that the partial graphs are generated by randomly erasing a fraction of the relations, which they are told in our experiments.

[3] Of course, we also need to assume that the chain is ergodic; a fair assumption given humans' non-zero probability of doing something strange/unexpected.

[4] As determining convergence can be non-trivial in certain cases, especially when the state space is large.



(a) (b)

Fig. 2. *Priors can be more precisely recovered by leveraging the MCMCP assumptions.* **(a)** We simulated data from our MCMCP Bayesian model on 5 nodes (34 non-isomorphic graphs), with a prior chosen to give an asymptotic mixing time of $\tau_m \sim 13$ iterations. Each simulation has 2048 data points, split into different chain lengths. We then fit a multinomial prior to these data, and measured the KL divergence from the true prior (that generated the data) for the fitted prior and for the sampled frequency of graphs. Error bars denote ±1 standard deviation. Sampling i.i.d. from the true prior is shown in gray as reference. Notice that using the graph frequency is doomed to fail when the chain length is short as there is not enough time for the chain to approach the prior. Moreover, fitting the prior does better than using the graph frequency even when the chain is much longer than τ_m. **(b)** Here we fix the chain length to 16, and vary the number of chains (same specifications as before). As the number of data points increases, fitting the prior continues to improve, while using the graph frequency asymptotes to a finite error.

3.2 A Natural Low-Dimensional Parametrization of Distributions Over Graphs

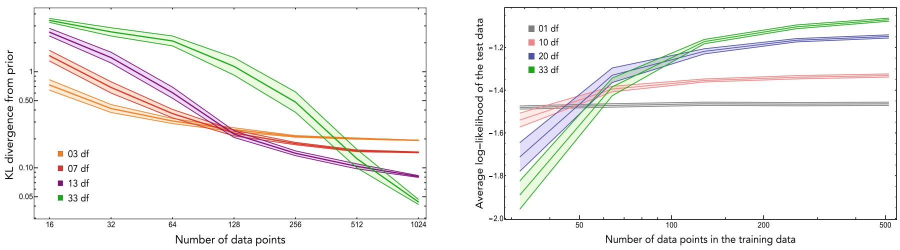
The number of non-isomorphic graphs $G(n)$ on n nodes grows superexponentially [15]; given limited data, even for moderate n we cannot sufficiently sample them all. For these cases, to obtain informative priors, we need to extend the probabilities to graphs that were not sampled. Our approach is to find a natural low-dimensional parameterization of the prior. Specifically, we propose to use the following form for the prior $\underline{\mathbf{p}}$:

$$\underline{\mathbf{p}} \propto \mathrm{ER}(1/2) * \exp \sum_{b=2}^{G(n)} c_b \underline{\mathbf{v}}_b \tag{2}$$

where c_b are the coefficients to be fit, $\underline{\mathbf{v}}_b$ is the b^{th} right eigenvector (ordered by decreasing eigenvalues) of the transition matrix $\underline{\underline{\mathbf{T}}}$ from Eq. 1, with the data generated by obscuring one relation of the underlying graph, and using an ER(1/2) (Erdős-Rényi model with $p = 1/2$) distribution as the prior.

This choice has several interesting properties,[5] for example, when $c_{i>2} = 0$, there is a unique correspondence between $c_2 \in (-\infty, \infty)$ and an $ER(p)$ prior with

[5] Formal details about this parametrization and its extension to hypergraphs to appear in a paper under preparation by Bravo Hermsdorff and Gunderson.



(a) (b)

Fig. 3. *Benefits of our low-dimensional smooth parametrization of the prior in the limited data regime.* Error bars denote ±1 standard error (SE). **(a)** *Improved accuracy.* We simulated data using the same specifications as in Fig. 2b, and fit the prior for several values of df. Notice that when data are limited, using a low-dimensional parametrization recovers the prior more accurately, but as the number of data points increases, using the full multinomial ($df = G(n) - 1$) does best. **(b)** *Better generalization.* We used 1210 data points from a single social cover story on 5 nodes. We randomly split them into test (698 data points) and training data, and fit the prior for several values of df. Notice that, in accord with the bias-variance tradeoff, using a low-dimensional parameterization for the prior results in better generalization (higher log-likelihood of the unseen data) when data are scarce, but as the number of data points increases, using the full multinomial does best.

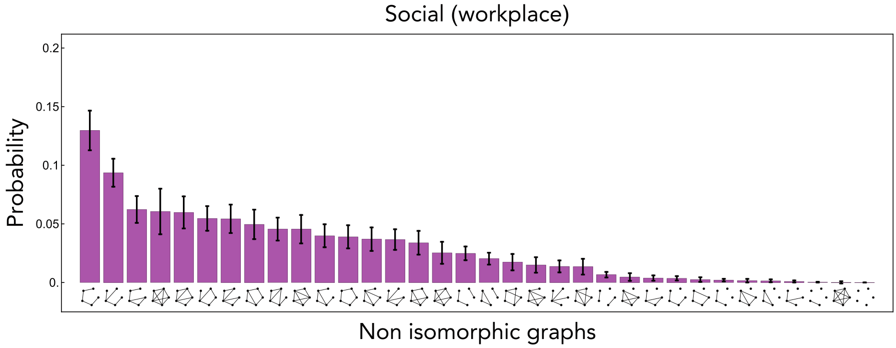
$p \in (0, 1)$. "Smoother" priors, parameterized by the number of degrees of freedom $1 \leq df \leq G(n) - 1$, are obtained by including only the longer-decaying modes (i.e., $c_{i>df+1} = 0$). In Fig. 3a, we show the effect of df when applying our model to simulated data. When all the graphs are sufficiently sampled, $df = G(n) - 1$ (equivalent to a full multinomial model) recovers the prior more accurately. In the limited data regime, however, using fewer coefficients does better, as it avoids overfitting. Figure 3b illustrates the associated improved generalization using human data.

3.3 Subjects' Priors Have Non-trivial Graphical Structure

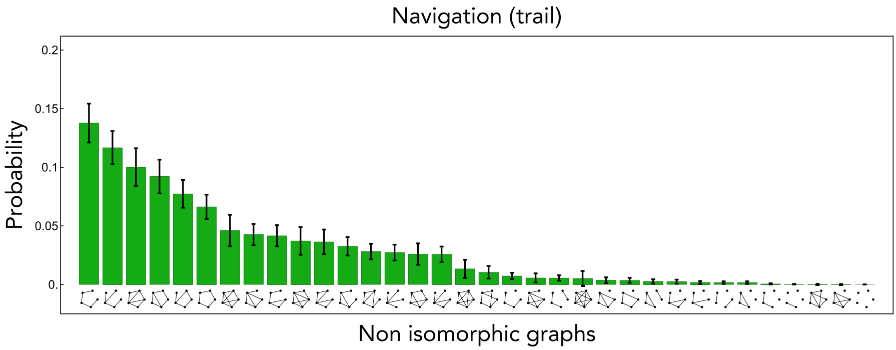
We compared several models for the priors using leave-p-out cross-validation (CV) [1] on data from our online experiments.[6] In particular, we considered several choices of df for our smooth parametrization of the prior, a full multinomial prior, and two "null models" (in the sense that they are not sensitive to graphical structure): *(1)* Erdős-Rényi (ER) prior, where the edges probabilities are independent and identical random variables; and *(2)* Prior over average degree, where the edge probabilities are no longer independent but are completely exchangeable (equivalent to only counting the *number* of edges).

For all priors we considered (different cover stories and numbers of nodes), a smooth parametrization of the prior did better (higher average log-likelihood in

[6] Our experiments use graphs on 4 to 10 nodes.



(a)



(b)

Fig. 4. *Deviations from basic null models suggests that the priors have non-trivial graphical structure.* To generate error bars, we simulated our Bayesian MCMCP model using the same set of partial graphs and the prior fitted to human data, and then fit our model to these simulated data. The error bars denote ±1 standard deviation from the mean fitted prior in the simulated data. **(a)** *Social:* Using 481 data points of a cover story of inferring the friendship network in a workplace with 5 people, we performed leave-10-out cross validation (CV). The figure displays the fit with the highest CV score (defined as the average log-likelihood per trial in the test sets). This fitted prior (31 *df*, CV score: −2.52, SE: 0.062) deviates significantly (t-test p-value: < .0001) from both null models (ER model, CV score: −2.79, SE: 0.043; and prior over average degree, CV score: −2.71, SE: 0.056). **(b)** *Navigation:* We performed the same analysis on a dataset (537 data points) of a cover story of inferring the trail map of a nature park with 5 sights. As before, the best fitted prior has 31 *df* (CV score: −2.41, SE: 0.066), and deviates significantly from both null models (ER model, CV score: −2.79, SE: 0.049; and prior over average degree, CV score: −2.71, SE: 0.061).

the CV test sets) than the two null models. Additionally, in all cases, the best fit had a relatively high number of *df*, suggesting that there is non-trivial graphical structure in subjects' priors. Figure 4 displays results for priors over 5 nodes. Moreover, as illustrated in Fig. 5 when comparing subjects' priors over different

Training \ Test	Soc. Class	Soc. Work	Nav. City	Nav. Trail
Soc. Class	0.953(3)	0.992(2)	0.957(2)	0.947(2)
Soc. Work	0.974(2)	0.979(3)	0.896(2)	0.918(2)
Nav. City	0.925(3)	0.886(3)	0.975(3)	0.985(2)
Nav. Trail	0.852(5)	0.851(4)	0.946(4)	0.964(4)

Fig. 5. *Domain-dependent priors.* We compared the priors of 4 experiments with different cover stories in 5 nodes by randomly splitting these 4 datasets in training (382 data points) and test (96 data points) data, and fitting a full multinomial model to each training set individually as well as to the aggregated 4 training sets. For each test set, we calculated the likelihood ratio (LR) between the individual fits and the aggregated fit. Notice that the LR is larger when the training and test data share the same domain (as indicate by the block diagonal structure of the table). Moreover, the fact that all LR are smaller than 1 indicates that we still need to collect more data.

cover stories on the same number of nodes, the priors between the navigation and social domains were notably different, while showing fewer differences between different contexts in the same domain (navigation: city and nature park; social: coworkers and students). This raises the interesting possibility that priors over task graphs are sensitive to the more abstract structure of a task (domain) rather than its specific context, which would allow for broader generalization.

3.4 General Framework for Quantifying Priors over Exchangeable Random Structures

Finally, we extend our results to the more general case of quantifying priors over exchangeable random structures [14], where the partial data are generated by randomly obscuring a given fraction of the sequence. The relevant parameters are: $\mathcal{A}$, the alphabet; ℓ, the string length; m, the number of relations obscured; and $\mathcal{G}$, the group under which the sequence is exchangeable. For example, for unordered binary strings, $\mathcal{A} = \{0, 1\}$ and $\mathcal{G}$ is the *full* permutation group $\mathcal{S}_\ell$ acting on the entries of strings of length ℓ; for simple graphs, the binary string is length $\ell = \binom{n}{2}$ and $\mathcal{G}$ is the permutation group $\mathcal{S}_n$, where n is the number of *nodes*. An element in $\mathcal{G}$ induces a permutation of the indices $\{1, \ldots, \ell\}$, and thus a permutation of the elements in $\mathcal{A}^\ell$. This action of $\mathcal{G}$ induces an equivalence relation on $\mathcal{A}^\ell$ ($x \sim y$ if $\exists g \in \mathcal{G}$ s.t. $x = g.y$), partitioning it into equivalence classes. For example, for unordered binary strings, they are partitioned into $\ell+1$ sets, one for each possible sum $(0, ..., \ell)$; for simple graphs, the partitions correspond to the non-isomorphic graphs on n nodes. The condition of exchangeability under $\mathcal{G}$ means that probabilities are assigned to these partitions, with elements in the same partition having equal probability.

4 Discussion

In this work, we develop a formal framework to quantify humans' priors over exchangeable random structures, and apply it to the case of non-isomorphic graphs representing the structure of several navigation and social tasks (instantiated using our new online experimental platform). We believe that navigation and social interaction tasks are a good starting point as they are an integral part of humans' lives and their different structure can provide interesting comparisons. Although we are currently collecting more data to improve our statistics, and doing further detailed analysis of the data (e.g., modeling priors over larger number of nodes, and building generative models that explains subjects' priors), the results reported here are encouraging, in the sense that there appears to be non-trivial domain-dependent structure in subjects' priors.

It is important to highlight that our model makes several assumptions about subjects' behavior that could not hold in practice (e.g., that they are "Bayesian"), and that are hard to test empirically (due, e.g., to interactions among the effects of the different assumptions). We argue that for our method to shed light on our understanding of our priors over tasks structure, it is more important to establish whether the priors we obtain from our experiments are in fact "meaningful" (in the sense of being used in practice) than whether the data violate certain assumptions or not.[7] Hence, to test the behavioral relevance of our measured priors, we are beginning to test if subjects learn faster/have better performance in experimental tasks (distinct from our MCMCP experiments) with structures consistent with these priors. Additionally, we are working on developing principled ways to test if analogous real-world datasets have structure similar to the priors.

In sensory and motor neuroscience, quantifying humans' priors over these systems has led to significant insights [7,8,12,13]. For example, several visual illusions can be understood as resulting from priors that encode the structure of naturalistic scenes [10,11]. Analogously, understanding our beliefs about the structure of new tasks could lead to a deeper understanding of our learning and generalization abilities (as well as their failure modes) [2,4]. We hope that the approach we proposed will pave the way towards the more ambitious goals of formalizing what a general "task" is, and of providing unifying principles that explain what makes a given task easy or hard for a human to solve.

References

1. Arlot, S., Alan, C.: A survey of cross-validation procedures for model selection. Stat. Surv. **4**, 40–79 (2010). https://doi.org/10.1214/09-SS054
2. Bengio, Y., Courville, A., Vincent, P.: Representation learning: a review and new perspectives. IEEE Trans. Pattern Anal. Mach. Intell. **35**, 1798–1828 (2013). https://doi.org/10.1109/TPAMI.2013.50
3. Bellman, R.E.: Dynamic Programming. Princeton University Press, Princeton (1957)

[7] Science is full of examples of reduced models (thermodynamics, effective field theories, magnetohydrodynamics, etc.) that provide powerful predictions even when applied to systems that violate their assumptions.

4. Botvinick, M., Weinstein, A., Solway, A., Barto, A.: Reinforcement learning, efficient coding, and the statistics of natural tasks. Curr. Opin. Behav. Sci. **5**, 71–77 (2015). https://doi.org/10.1016/j.cobeha.2015.08.009
5. Brady, T.F., Konkle, T., Alvarez, G.A.: Compression in visual working memory: using statistical regularities to form more efficient memory representations. J. Exp. Psychol. Gen. **138**, 487–502 (2009). https://doi.org/10.1037/a0016797
6. Canini, K.R., Griffiths, T.L., Vanpaemel, W., Kalish, M.L.: Revealing human inductive biases for category learning by simulating cultural transmission. Psychon. Bull. Rev. **21**, 785–793 (2014). https://doi.org/10.3758/s13423-013-0556-3
7. Field, D.F.: What the statistics of natural images tell us about visual coding. Proc. SPIE Int. Soc. Opt. Eng. **1077**, 269–276 (1989). https://doi.org/10.1117/12.952724
8. Graziano, M.S.A.: Cortical action representations. In: Toga, A.W., Poldrack, R.A. (eds.) Brain Mapping: An Encyclopedic Reference. Elsevier, Amsterdam (2014). http://www.princeton.edu/~graziano/Graziano_encyclopedia_2015.pdf
9. Griffiths, T., Kalish, M.: Language evolution by iterated learning with Bayesian agents. Cogn. Sci. **31**, 441–480 (2007). https://doi.org/10.1080/15326900701326576
10. Howe, C.Q., Purves, D.: The Müller-Lyer illusion explained by the statistics of image source relationships. Proc. Natl. Acad. Sci. **102**(4), 1234–1239 (2005). https://doi.org/10.1073/pnas.0409314102
11. Howe, C.Q., Yang, Z., Purves, D.: The Poggendorff illusion explained by natural scene geometry. Proc. Natl. Acad. Sci. **102**(21), 7707–7712 (2005). https://doi.org/10.1073/pnas.0502893102
12. Lewicki, M.S.: Efficient coding of natural sounds. Nat. Neurosci. **5**(4), 356–363 (2002). https://doi.org/10.1038/nn831
13. Orbán, G., Fiser, J., Aslin, R.N., Lengyel, M.: Bayesian learning of visual chunks by human observers. Proc. Natl. Acad. Sci. **105**(7), 2745–2750 (2008). https://doi.org/10.1073/pnas.0708424105
14. Orbanz, P., Roy, D.M.: Bayesian models of graphs, arrays and other exchangeable random structures. IEEE Trans. Pattern Anal. Mach. Intell. **37**, 437–461 (2015). https://doi.org/10.1109/TPAMI.2014.2334607
15. The on-line encyclopedia of integer sequences (OEIS). https://oeis.org/A000088

Mining the Temporal Structure of Thought from Text

Mei Mei(✉), Zhaowei Ren, and Ali A. Minai

Department of Electrical Engineering and Computer Science,
University of Cincinnati, Cincinnati, OH 45221-0030, USA
meimi@mail.uc.edu, rzwdsg@gmail.com, Ali.Minai@uc.edu

Abstract. Thinking is a self-organized dynamical process and, as such, interesting to characterize. However, direct, real-time access to thought at the semantic level is still very limited. The best that can be done is to look at spoken or written expression. The question we address in this research is the following: *Is there a characteristic pitch of thought*? To begin answering this complex question, we look at text documents from several large corpora at the sentence level – i.e., using sentences as the units of meaning – and considering each document to be the result of a random process in semantic space. Given a large corpus of multi-sentence documents, we build a lexical association network representing associations between words in the corpus. This network is used to induce a semantic similarity metric between sentences, and each document is segmented into multi-sentence *semantically coherent blocks* (SCBs) with occasional connecting text between the blocks. Based on this segmentation, the process of document generation is modeled as a sticky Markov chain at the sentence level. We show that most documents across all the corpora are sequences of blocks with a very consistent mean length of 6.4 sentences across the corpora. This consistency suggests that a value of 6-7 sentences may be the typical mean length for single coherent thoughts in texts. We have also described several ways of visualizing the semantic structure of documents in space and time.

Keywords: Semantic dynamics · Text analysis · Text segmentation

1 Introduction

Studying and characterizing thoughts is an interesting, yet difficult task since thinking is a complex process which cannot be observed or measured directly. In the past decades, some methodologies have been developed in various research fields [4], including neuroscience [3,13,15], psychology [10] and artificial intelligence [6]. As thoughts cannot be measured directly, one possibility is to look at writings or speeches, which can be seen as partial, externalized and concrete representations of the underlying thought process in the mind of the writer or speaker. As a first step towards this, we report on a method for analyzing the temporal semantic structure of texts and characterizing this structure.

A. J. Morales et al. (Eds.): ICCS 2018, SPCOM, pp. 291–298, 2018.
https://doi.org/10.1007/978-3-319-96661-8_31

Semantic analysis of text corpora is an issue of broad interest, and has been addressed in terms of text representation [1,14], topic modeling [2] and text segmentation, [7,9,11]. Our method is based on segmenting documents into sequences of *semantically coherent blocks* (SCBs) and *gaps*, and then characterizing the sequences using a Markovian model. We also demonstrate several possible signatures for visualizing the temporal semantic structure in individual documents.

2 Document Segmentation

Our approach assumes that the documents being analyzed belong to a large and topically coherent corpus, so that patterns of word usage within the corpus can be used to infer semantic similarity. The results in this report are based on a corpus comprising papers from the *Proceedings of International Joint Conference on Neural Networks* (IJCNN) from three different years.

The process of segmentation and document characterization involves the several stages as described in this section.

1. Preprocessing: The raw text corpus is first processed to remove any superfluous material (page numbers, figures, etc.). All words are stemmed using a Porter stemmer. Common words such as prepositions, articles and conjunctions are removed along with some other corpus-specific *stop words* identified using methods we have described elsewhere [8].

2. Construction of the Lexical Network: The corpus is analyzed at the level of words and sentences to build a *lexical association network* (LAN). Each node in this network is a unique word from the corpus vocabulary, and the weighted edges between word pairs represent the tendency of the words to co-occur in the same sentence.

3. Calculation of Sentence Similarities: For each document in the corpus, a similarity value is calculated for each pair of sentences based on how strongly the words of each are connected with the words of the other in the LAN. Thus, sentences that use highly correlated words can be similar even if they share no words. This step gives a *sentence similarity matrix* (SSM) for each document.

4. Initial Segmentation: The SSM for each document is processed to detect semantically coherent blocks, where each block is defined as a set of consecutive sentences with high mutual similarities. Detecting semantic blocks is a non-trivial problem since SSMs are usually noisy. We use the following edge-detection heuristic to find blocks: (a) The mean similarity of each sentence with its k previous sentences and k subsequent sentences is calculated; (b) Potential block end and start points are defined by detecting sharp drops and rises of those two values, respectively; (c) Beginning with the first 'start' point, alternating 'start' and 'end' points are identified to mark the blocks. Blocks are constrained to be of length 3 or more. Sentence sequences between successive blocks are designated as gaps. Thus, each document gets segmented into a sequence of blocks and gaps termed its *initial segmentation.*

5. Refinement of Segmentation: Since the initial segmentation is based on a heuristic, it can be quite inaccurate. We further refine it by merging and expanding blocks, and removing low quality blocks – all based on a *block coherence* metric calculated for each block as the mean similarity of all its sentence pairs.

Figure 1 shows an example of the SSM and the final segmentation for a sample document.

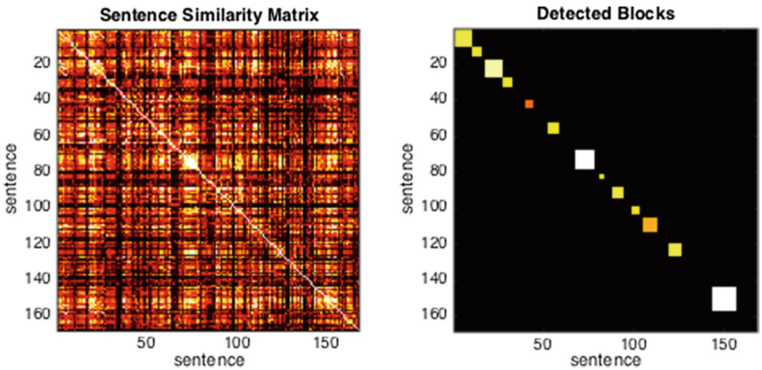


Fig. 1. Left: Sentence similarity matrix of a document. Right: Segmentation of the document. The colors of the blocks represent the quality (coherence), with a lighter color indicating higher quality.

3 Markov Model

After each document is segmented into a sequence of blocks and gaps, its semantic structure is modeled as being characterized by a Markov chain, with different types of blocks and gaps as states. Since a block can be followed by another block or a gap, but a non-terminal gap can only be followed by a block, it is useful to keep track of the number of consecutive blocks before transitioning to a gap. This is done by defining the *block order*: A block has order d if it is the dth consecutive block since the last gap. Thus, a block that follows after a gap is defined as a *first order* block, denoted by B_1. A block that follows directly after B_1 is defined as a *second order* block, denoted by B_2, and so on. Analysis of the documents in the IJCNN corpus indicates that blocks of order of 6 or higher are too infrequent to provide good statistics, so they are treated as a single class, B_6^+. This leads to a 7-state discrete-time Markov model, with six block states, B_1 through B_6^+, and the gap state, G, as shown in Fig. 2 (left), with each sentence representing a step. However, since each block is constrained to be at least three sentences long, each block state actually consists of three microstates as shown in Fig. 2 (inset). Once the process enters a block state, B_k in microstate a at its initial sentence, it must transition through microstate b to microstate c over the next two consecutive sentences, after which it can either remain in B_k, or transition to the next block state or G.

To validate the model by comparing the distribution of block lengths and frequencies predicted by the model against actual data. If the block transition

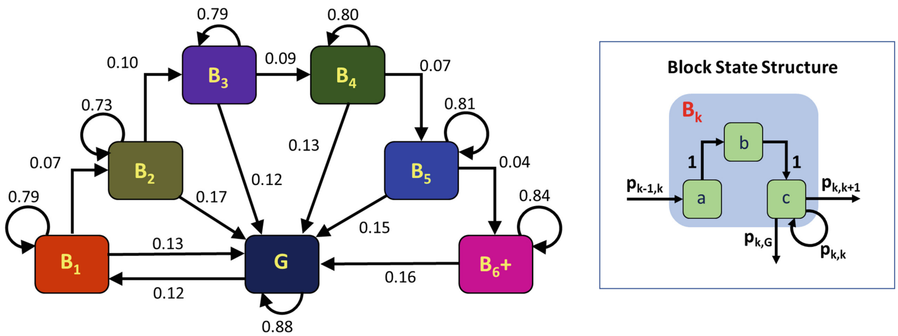


Fig. 2. Left: 7-state Markov model with transition probabilities estimated based on the data from the IJCNN corpus. Right (Inset): Internal microstructure of block states.

process can be modeled by Markov process, the probability of block length $2+k$ should satisfy a goemetric distribution, $P(k+2) = (1-q)q^k \sim q^k$, where q is the probability of staying within current block for any sentences after the first two in the block. Thus, the semi-log plot of the histogram of block length should be a straight line. Similar calculation can be applied on gap analysis.

Plotting the histogram of block length and gap length from experimental data on a semi-log scale (not shown) indicates that the block length curve is almost linear, which validates the Markov model. The probability q can be estimated by looking at the probability of blocks of length k and $k+d$, where $k > 2$.

$$\hat{q} = \sqrt[d]{\frac{P(k+d)}{P(k)}} \approx \sqrt[d]{\frac{N_{k+d}}{N_k}} \tag{1}$$

where N_j is the number of blocks of length j.

The estimate $\hat{q}$ is obtained by taking two points from the experimental data, $(k = 7, N_k = 276)$ and $(k+l = 16, N_{k+l} = 30)$, which gives $\hat{q} \approx \sqrt[16-7]{\frac{30}{276}} \approx 0.78$. The result is quite consistent with experimental data shown in Fig. 2. Thus, we conclude the data is fit well by the proposed Markov chain model.

The average block length is also calculated for each year of the IJCNN corpus, giving an approximate average block length of 6.4 with a standard deviation of 4.2 for every corpus. This consistency provides a characterization of pitch of thought, showing that, at least for the process underlying the composition of documents in the IJCNN corpus, a coherent theme typically lasts about 6 or 7 sentences, albeit with a significant amount of variation.

4 Block Embedding and Clustering

After all the documents in the corpus are segmented into blocks and gaps, it is interesting to cluster them based on their semantic similarity. This, in turn, requires that every block be embedded in a feature space. The most obvious feature space is provided by the words in the corpus' vocabulary. In principle, a block can be represented by a vector with each component equal to the number of occurrences of the corresponding word in the block. However, given that blocks are typically just a few sentences long and the corpus vocabulary can have several thousand words, the resulting representations are extremely sparse and high-dimensional, which makes them difficult to cluster meaningfully. Instead, we adopt an alternative approach where a *semantic embedding space* (SES) is defined by extracting a small number of *topics* from the entire corpus, and then representing each block as a point in this vector space with its coordinates given by the block's consistency with each topic. While this *topic space* is useful for clustering, it is still too high-dimensional to visualize. For this, we apply a further dimensionality reduction method to map all blocks into a 2-dimensional *reduced embedding space* (RES) as described below. This allows each document to be visualized as a trajectory of blocks in the RES.

Topic Extraction: To extract topics from the corpus, we use the latent Dirichlet allocation (LDA) algorithm [2]. LDA models each document as a mixture of topics, $z_1, ..., z_T$, where each topic, z_j, defines a multinomial distribution θ_j over the vocabulary. The topics are obtained through Gibbs sampling of the corpus, and result in the generative assignment of topic strengths to each document. While LDA does not provide a way to determine the optimal number of topics, we determined empirically that $T = 25$ was a good choice for the corpus.

Block Embedding in Topic Space: Once topics have been generated, the words in each block are used to determine that block's consistency with each topic, and these values give a T-dimensional topic space vector representation for each block. It should be noted that, while the semantic content of each topic is of interest, it is not a significant issue for the use of topics to define a feature space for clustering as long as the topics are sufficiently distinct from each other.

Block Clustering: The topic space representations of blocks are clustered using the standard k-means clustering algorithm. The silhouette metric [12] is used as the criterion for determining the optimal number of clusters. Figure 3 shows the clustered blocks in topic space from a subset of the corpus. As shown in the figure, most clusters are clear and homogeneous, documents get clustered together when they share one or two strong topics. However, a few clusters are more diffuse, and represent blocks with complex semantics.

Visualization of Document Signatures and Trajectories: After each block in the corpus is assigned to a cluster, a document can be represented using a strip signature in two ways: (1) A block-level signature where a strip of blocks and gaps in the document are colored based on the cluster assignment of each block; and (2) A *topic-level signature*, which is a *topic* × *sentence* matrix heat-map, with the intensity of cell (j, k) indicating the consistency of the kth sentence in the document with topic z_j. Figure 4 shows both signatures for two documents. It can be seen from the block-level signatures that the document on the left is confined mainly to one cluster, with a brief digression to another one. The document on the right, in contrast, ranges over several clusters, indicating that

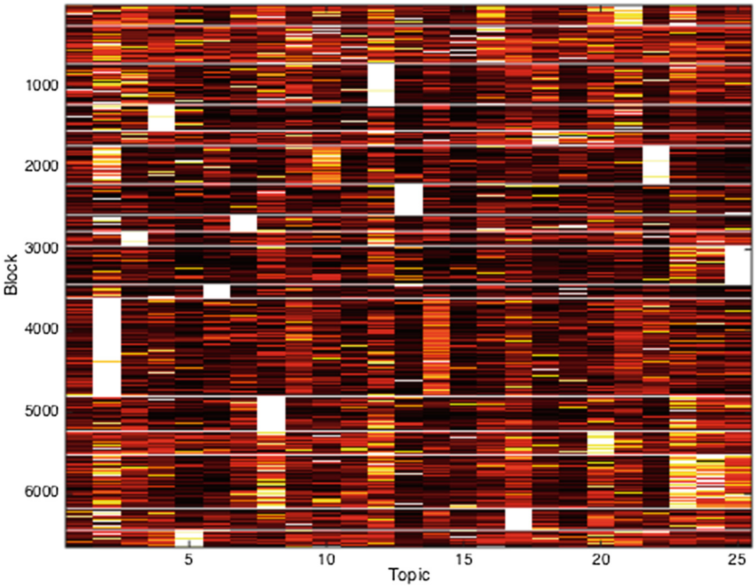


Fig. 3. Topic-level representation of 6,656 blocks in a corpus, sorted into 17 clusters. White solid lines indicate the cluster boundaries.

it is semantically a more complex document than the first one. The topic-level signature provides a more detailed snapshot of how topics are distributed in the document. Together, the two signatures are "barcode" style snapshots of a document's semantic structure.

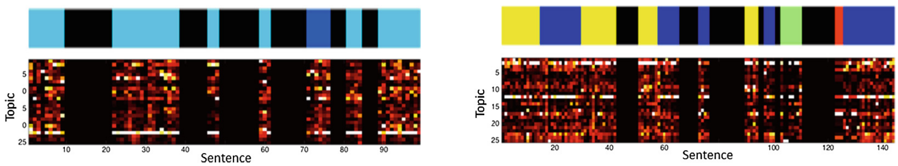


Fig. 4. Block-level and topic-level signatures for two documents. Left: Concentrated document focused mostly in one cluster. Right: Diffuse document that traverses multiple clusters.

It is also interesting to visualize the temporal dynamics of individual documents in semantic space. To do this, we use a dimensionality-reduction approach called *stochastic nearest neighbors embedding* (SNE) [5] to map the T-dimensional topic-space representation to a 2-dimensional one. Each block is then plotted in this space using its cluster color, and successive blocks (ignoring gaps) are connected by lines to show the document's semantic trajectory reflecting the temporal structure of meaning in the minds of the authors. Figure 5 shows the trajectories for the same concentrated and diffuse documents whose signatures were shown in Fig. 4. The former has a tight trajectory with one distant excursion, while the latter has a more wide-ranging trajectory.

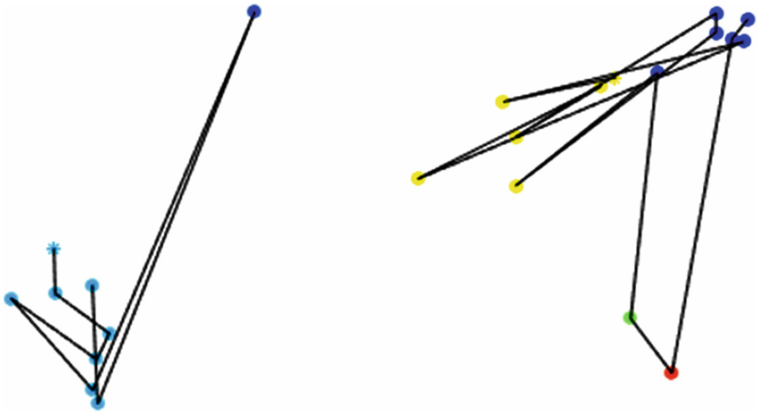


Fig. 5. Block trajectories for the two documents from Fig. 4 in the reduced embedding space. Left: Concentrated document. Right: Diffuse document.

5 Conclusion

Based on several datasets of research publications, our results indicate that, on average, people tend to write 6 to 7 consecutive semantically coherent sentences before shifting to a new block or a gap, and that the data across a large number of documents is fit quite well by a Markov process under the stipulation that a semantic block must not be less than three consecutive sentences in length. These results are consistent across all tested corpora, though this conjecture must still be verified on more diverse corpora. We have also demonstrated several possible signatures for visualizing the temporal semantic structure in individual documents.

Acknowledgement. This work was supported in part by National Science Foundation INSPIRE grant BCS-1247971 to Ali Minai.

References

1. Aggarwal, C.C., Zhao, P.: Towards graphical models for text processing. Knowl. Inform. Syst. **36**(1), 1–21 (2013). https://doi.org/10.1007/s10115-012-0552-3
2. Blei, D.M., Ng, A.Y., Jordan, M.I.: Latent Dirichlet allocation. J. Mach. Learn. Res. **3**, 993–1022 (2003). https://doi.org/10.1162/jmlr.2003.3.4-5.993
3. Canolty, R.T., Soltani, M., Dalal, S.S., Edwards, E., Dronkers, N.F., Nagarajan, S.S., Kirsch, H.E., Barbaro, N.M., Knight, R.T.: Spatiotemporal dynamics of word processing in the human brain. Front. Neurosci. **1**(1), 185–196 (2007). https://doi.org/10.3389/neuro.01.1.1.014.2007

4. Friedenberg, J., Silverman, G.: Introduction: exploring inner space. In: Brace-Thompson, J., Crouppen, M.B., Robinson, S. (eds.) Cognitive Science An Introduction to the Study of Mind, chapter 1, pp. 2–3. Sage Publications, Inc., Thousand Oaks (2006)
5. Hinton, G.E., Roweis, S.T.: Stochastic neighbor embedding. In: Advances in neural information processing systems, pp. 833–840 (2002)
6. Hogan, J.P.: Mind Matters: Exploring the World of Artificial Intelligence, 1st edn. Ballantine Publication Group, New York (1998)
7. Lamprier, S., Amghar, T., Levrat, B., Saubion, F.: SegGen: A genetic algorithm for linear text segmentation. In: Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI), pp. 1647–1652 (2007)
8. Mei, M., Vanarase, A., Minai, A.A.: Chunks of thought: finding salient semantic structures in texts. In: Proceedings of IJCNN 2014 (2014)
9. Misra, H., Yvon, F., Cappé, O., Jose, J.: Text segmentation: a topic modeling perspective. Inform. Process. Manag. **47**(4), 528–544 (2011). https://doi.org/10.1016/j.ipm.2010.11.008
10. Morewedge, C.K., Giblin, C.E., Norton, M.I.: The (perceived) meaning of spontaneous thoughts. J. Exp. Psychol. Gen. **143**(4), 1742–1754 (2014). https://doi.org/10.1037/a0036775
11. Riedl, M., Biemann, C.: Text segmentation with topic models. J. Lang. Technol. Comput. Linguist. **27**(1), 47–69 (2012)
12. Rousseeuw, P.J.: Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. Comput. Appl. Math. **20**, 53–65 (1987)
13. Shen, G., Horikawa, T., Majima, K., Kamitani, Y.: Deep image reconstruction from human brain activity. bioRxiv (2017). 10.1101/240317
14. Turian, J., Ratinov, L., Bengio, Y.: Word representations: a simple and general method for semi-supervised learning. In: Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics, July, pp. 384–394 (2010)
15. Wang, J., Cherkassky, V.L., Just, M.A.: Predicting the brain activation pattern associated with the propositional content of a sentence: Modeling neural representations of events and states. Hum. Brain Mapp. **38**, 4865–4881 (2017). https://doi.org/10.1002/hbm.23692

(Anti)Fragility and Convex Responses in Medicine

Nassim Nicholas Taleb(✉)

Tandon School of Engineering, New York University, New York, NY 11201, USA
NNT1@nyu.edu

Abstract. This paper applies risk analysis to medical problems, through the properties of nonlinear responses (convex or concave). It shows (1) necessary relations between the nonlinearity of dose-response and the statistical properties of the outcomes, particularly the effect of the variance (i.e., the expected frequency of the various results and other properties such as their average and variations); (2) The description of "antifragility" as a mathematical property for local convex response and its generalization and the designation "fragility" as its opposite, locally concave; (3) necessary relations between dosage, severity of conditions, and iatrogenics.

Iatrogenics seen as the tail risk from a given intervention can be analyzed in a probabilistic decision-theoretic way, linking probability to nonlinearity of response. There is a necessary two-way mathematical relation between nonlinear response and the tail risk of a given intervention.

In short we propose a framework to integrate the necessary consequences of nonlinearities in evidence-based medicine and medical risk management.

Keywords: Evidence based medicine · Risk management
Nonlinear responses

Comment on the Notations: we use x for the dose, $S(x)$ for the response function to x when is sigmoidal (or was generated by an equation that is sigmoidal), and $f(x)$ when it is not necessarily so.

1 Background

Consideration of the probabilistic dimension has been made explicitly in some domains, for instance there are a few papers linking Jensen's inequality and noise in pulmonary ventilators: papers such as Brewster et al. [1], Graham et al. [2], Funk [3], Arold et al. [4], Amato et al. [5]. In short, to synthesize the literature, continuous high pressures have been shown to be harmful with increased mortality, but occasional spikes of ventilation pressures can be advantageous with recruitment of collapsed alveoli, and do not cause further increased mortality. But explicit probabilistic formulations are missing in other domains, such

A. J. Morales et al. (Eds.): ICCS 2018, SPCOM, pp. 299–325, 2018.
https://doi.org/10.1007/978-3-319-96661-8_32

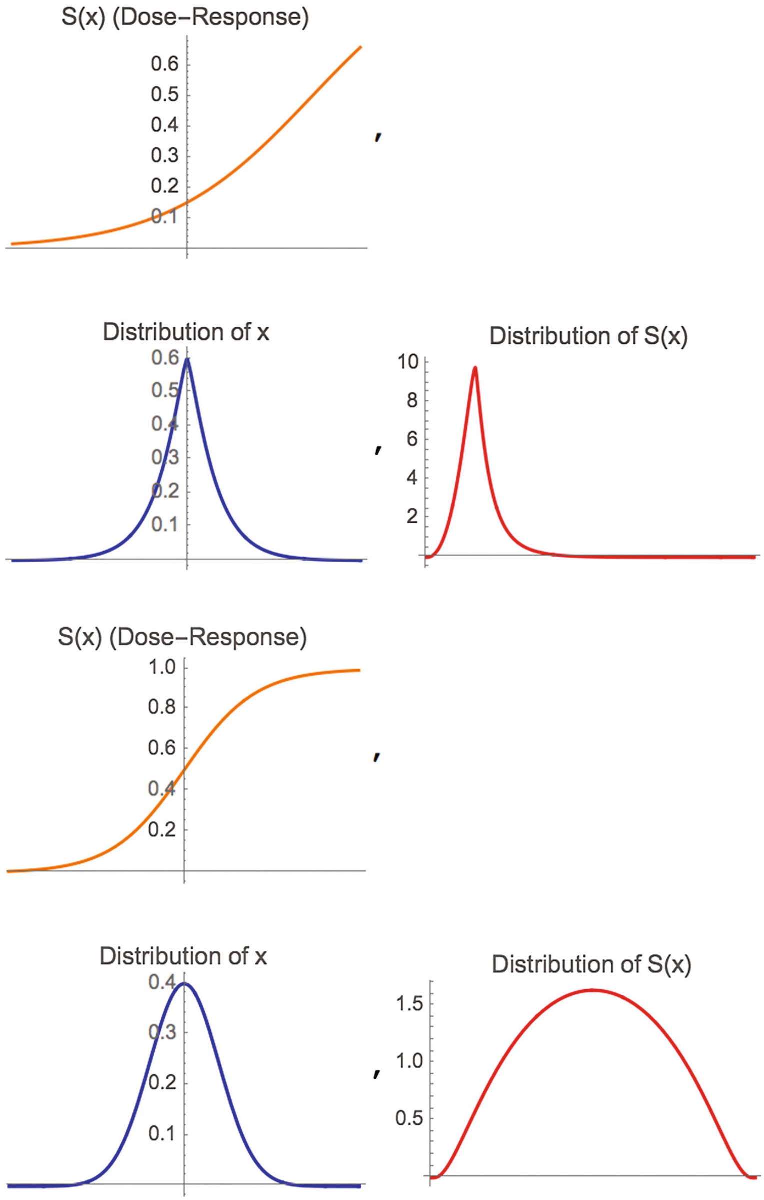


Fig. 1. These two graphs summarize the gist of this chapter: how we can go from the reaction or dose-response $S(x)$, combined with the probability distribution of x, to the probability distribution of $S(x)$ and its properties: mean, expected benefits or harm, variance of $S(x)$. Thus we can play with the different parameters affecting $S(x)$ and those affecting the probability distribution of x, to assess results from output. $S(x)$ as we can see can take different shapes (We start with $S(x)$ monotone convex (top) or the second order sigmoid).

as episodic energy deficit, intermittent fasting, variable uneven distribution of sub-groups (proteins and autophagy), vitamin absorption, high intensity training, fractional dosage, the comparative effects of chronic vs acute, moderate and distributed vs intense and concentrated, etc.

Further, the detection of convexity is still limited to local responses and does not appear to have led to decision-making under uncertainty and inferences on unseen risks based on the detection of nonlinearity in response, for example the relation between tumor size and the iatrogenics of intervention, or that between the numbers needed to treat and the side effects (visible and invisible) from an intervention such as statins or various blood pressure treatments.

The links we are investigating are mathematical and necessary. And they are two-way (work in both directions). To use a simple illustrative example:

- a convex response of humans to energy balance over a time window necessarily implies the benefits of intermittent fasting (seen as higher variance in the distribution of nutrients) over some range that time window,
- the presence of misfitness in populations that have exceedingly steady nutrients, and evidence of human fitness to an environment that provides high variations (within bounds) in the availability of food, both necessarily imply a nonlinear (concave) response to food over some range of intake and frequency (time window).

The point can be generalized in the same manner to energy deficits and the variance of the intensity of such deficits given a certain average.

Note the gist of our approach (Fig. 1): we are not asserting that the benefits of intermittent fasting or the existence of a convex response are true; we are just showing that if one is true then the other one is necessarily so, and building decision-making policies that bridge the two.

Finally, note that convexity in medicine is at two levels. First, understanding the effect of dosing and its nonlinearity. Second, at the level of risk analysis for patients.

1.1 Convexity and Its Effects

Let us define convexity as follows. Let the "response" function $f : \mathbb{R}^+ \to \mathbb{R}$ be a twice differentiable function. If over a range x $\in [a, b]$, over a set time period Δt, $\frac{\partial^2 f(x)}{\partial x^2} \geq 0$, or more practically (by relaxing the assumptions of differentiability), $\frac{1}{2}(f(x + \Delta x) + f(x - \Delta x)) \geq f(x)$, with $x + \Delta x$ and $x - \Delta x \in [a, b]$ then there are benefits or harm from the unevenness of distribution, pending whether f is defined as positive or favorable or modeled as a harm function (in which case one needs to reverse the sign for the interpretation).

In other words, in place of a dose x, one can give, say, 140% of x, then 60% of x, with a more favorable outcome one is in a zone that benefits from unevenness. Further, more unevenness is more beneficial: 140% followed by 60% produces better effects than, say, 120% followed by 80%.

We can generalize to comparing linear combinations: $\sum \alpha_i = 1$, $0 \leq |\alpha_i| \leq 1$, $\sum(\alpha_i f(x_i)) \geq f(\sum(\alpha_i x_i))$; thus we end up with situations where, for $x \leq b - \Delta$ and $n \in \mathbb{N}$, $f(nx) \geq n f(x)$. This last property describes a "stressor" as having higher intensity than zero: there may be no harm from $f(x)$ yet there will be one at higher levels of x.

Now if X is a random variable with support in $[a, b]$ and f is convex over the interval as per above, then

$$\mathbb{E}\left(f(x)\right) \geq f\left(\mathbb{E}(x)\right), \tag{1}$$

what is commonly known as Jensen's Inequality, see Jensen [6], Fig. 2. Further (without loss of generality), if its continuous distribution with density $\varphi(x)$ and support in $[a, b]$ belongs to the location scale family distribution, with $\varphi(\frac{x}{\sigma}) = \sigma\varphi(x)$ and $\sigma > 0$, then, with $\mathbb{E}_\sigma$ the indexing representing the expectation under a probability distribution indexed by the scale σ, we have:

$$\forall \sigma_2 > \sigma_1,\ \mathbb{E}_{\sigma_2}\left(f(x)\right) \geq \mathbb{E}_{\sigma_1}\left(f(x)\right) \tag{2}$$

The last property implies that the convexity effect increases the expectation operator. We can verify that since $\int_{f(a)}^{f(b)} y \frac{\phi\left(f^{(-1)}(y)\right)}{f'\left(f^{(-1)}(y)\right)} \mathrm{d}y$ is an increasing function of σ. A more simple approach (inspired from mathematical finance heuristics) is to consider for $0 \leq \delta_1 \leq \delta_2 \leq b - a$, where δ_1 and δ_2 are the mean expected deviations or, alternatively, the results of a simplified two-state system, each with probability $\frac{1}{2}$:

$$\frac{f(x-\delta_2) + f(x+\delta_2)}{2} \geq \frac{f(x-\delta_1) + f(x+\delta_1)}{2} \geq f(x) \tag{3}$$

This is of course a simplification here since dose response is rarely monotone in its nonlinearity, as we will see in later sections. But we can at least make claims in a certain interval $[a, b]$.

What Are We Measuring? Clearly, the dose (represented on the x line) is hardly ambiguous: any quantity can do, such as pressure, caloric deficit, pounds per square inch, temperature, etc.

The response, harm or benefits, $f(x)$ on the other hand, need to be equally precise, nothing vague, such as life expectancy differential, some index of health, and similar quantities. If one cannot express the response quantitatively, then such an analysis cannot apply.

1.2 Antifragility

We define as locally antifragile[1] a situation in which, over a specific interval $[a, b]$, either the expectation increases with the scale of the distribution as in Eq. 2, or

[1] The term antifragile was coined in Taleb [7] inspired from mathematical finance and derivatives trading, by which some payoff functions respond positively to increase in volatility and other measures of variation, a term in the vernacular called "long gamma".

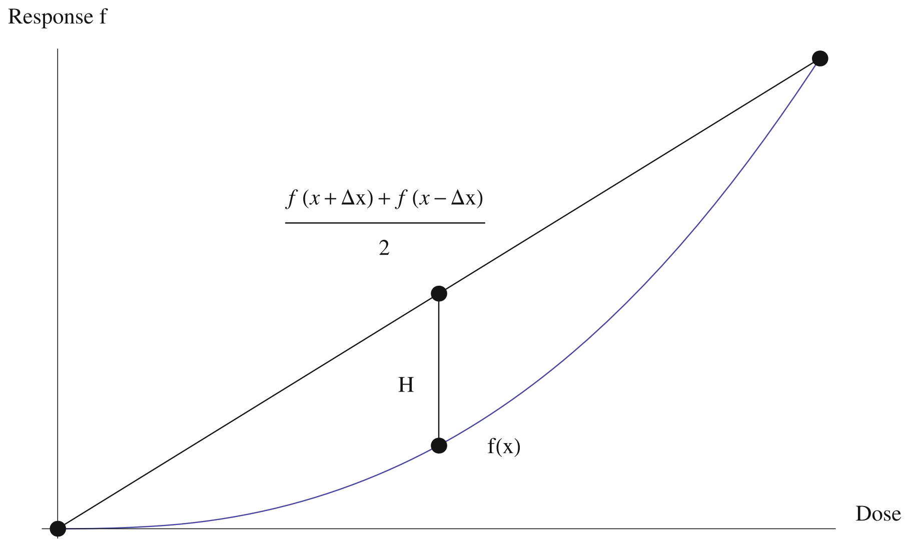


Fig. 2. Jensen's inequality

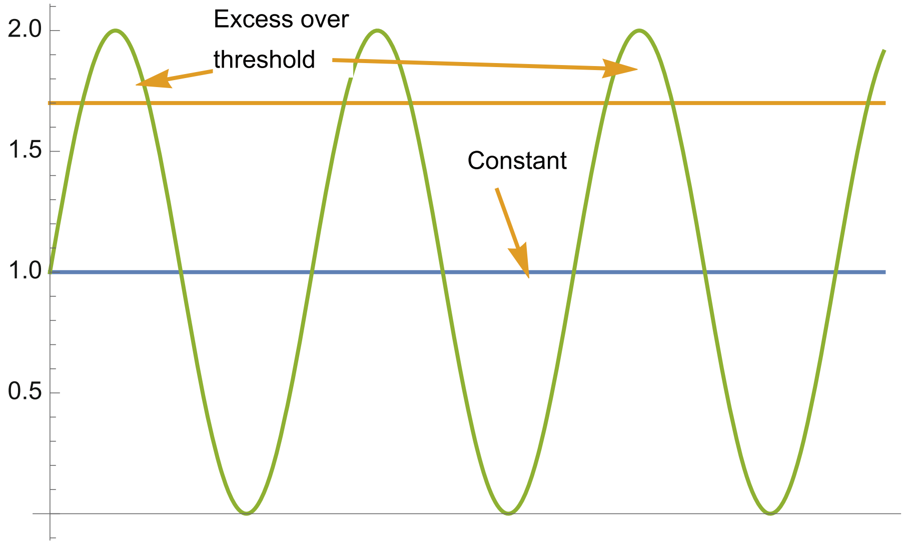


Fig. 3. The figure shows why fractional intervention can be more effective in exceeding a threshold than constant dosage. This effect is similar to **stochastic resonance** known in physics by which noise cause signals to rise above the threshold of detection. For instance, genetically modified BT crops produce a constant level of pesticide, which appears to be much less effective than occasional manual interventions to add doses to conventional plants. The same may apply to antibiotics, chemotherapy and radiation therapy.

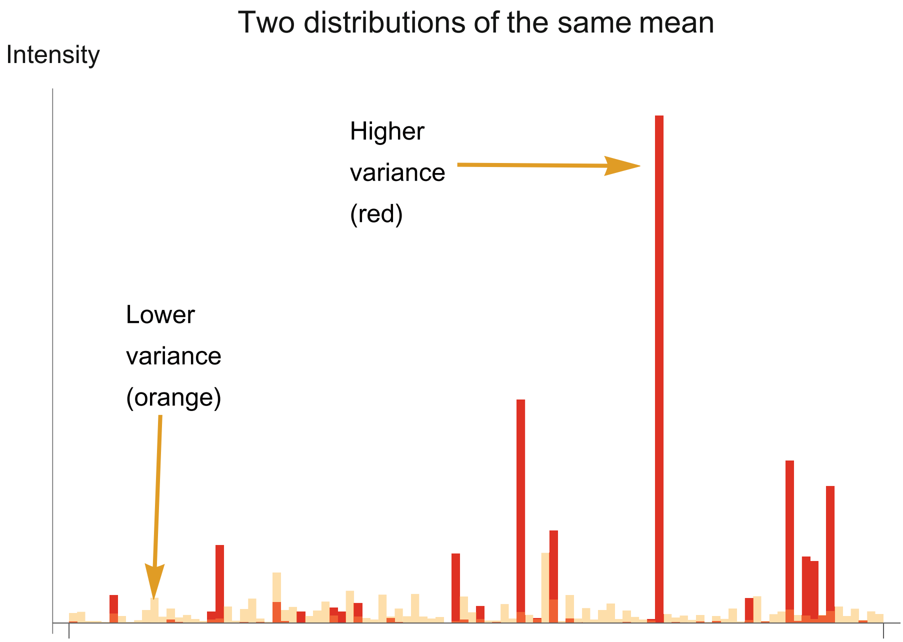


Fig. 4. An illustration of how a higher variance (hence scale), given the same mean, allow more spikes –hence an antifragile effect. We have a Monte Carlo simulations of two gamma distribution of same mean, different variances, $X_1 \sim G(1,1)$ and $X_2 \sim G(\frac{1}{10}, 10)$, showing higher spikes and maxima for X_2. The effect depends on norm $||.||_\infty$, more sensitive to tail events, even more than just the scale which is related to the norm $||.||_2$.

the dose response is convex over the same interval. The term in Taleb [7] was meant to describe such a situation with precision: any situation that benefits from an increase in randomness or variability (since σ, the scale of the distribution, represents both); it is meant to be more precise than the vague "resilient" and bundle behaviors that "like" variability or spikes. Figures 3, 4, 5 and 6 describe the threshold effect on the nonlinear response, and illustrates how they qualify as antifragile.

1.3 The First Order Sigmoid Curve

Define the sigmoid or sigmoidal function (Fig. 7) as having membership in a class of function $\mathfrak{S}$, $S : \mathbb{R} \to [L, H]$, with additional membership in the $\mathbb{C}^2$ class (twice differentiable), monotonic nonincreasing or nondecreasing, that is let $S'(x)$ be the first derivative with respect to x: $S'(x) \geq 0$ for all x or $S'(x) \leq 0$. We have:

$$S(x) = \begin{cases} H \text{ as } x \to +\infty; \\ L \text{ if } x \to -\infty. \end{cases},$$

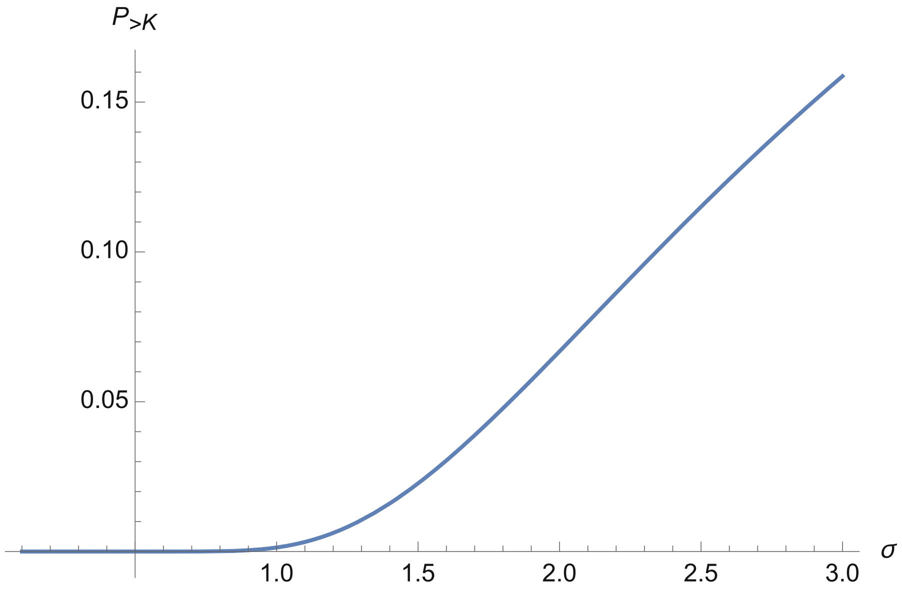


Fig. 5. Representation of Antifragility of Fig. 4 in distribution space: we show the probability of exceeding a certain threshold for a variable, as a function of σ the scale of the distribution, while keeping the mean constant.

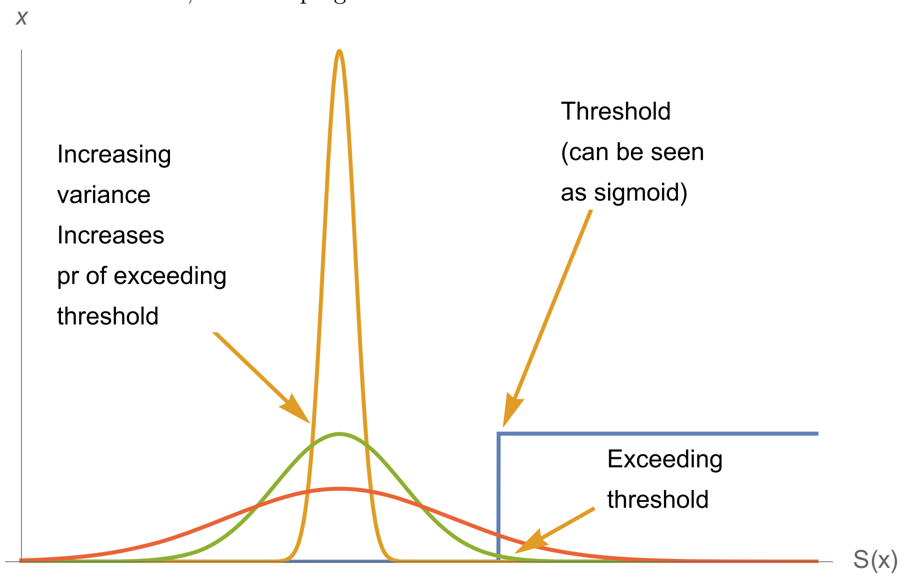


Fig. 6. How an increase in variance affects the threshold. If the threshold is above the mean, then we are in the presence of convexity and variance increases expected payoff more than changes in the mean, in proportion of the remoteness of the threshold. Note that the tails can be flipped (substituting the left for the right side) for the harm function if it is defined as negative.

which can of course be normalized with $H = 1$ and $L = 0$ if S is increasing, or vice versa, or alternatively $H = 0$ and $L = -1$ if S is increasing. We can define the simple (or first order) sigmoid curve as having equal convexity in one portion and concavity in another: $\exists k > 0$ s.t. $\forall x_1 < k$ and $x_2 > k$, $\operatorname{sgn}(S''(x_1)) = -\operatorname{sgn}(S''(x_2))$ if $|S''(x_2)| \geq 0$.

Now all functions starting at 0 will have three possible properties at inception, as in Fig. 8:

- concave
- linear
- convex

The point of our discussion is the latter becomes sigmoid and is of interest to us. Although few medical examples appear, under scrutiny, to belong to the first two cases, one cannot exclude them from analysis. We note that given that the inception of these curves is 0, no linear combination can be initially convex unless the curve is convex, which would not be the case if the start of the reaction is at level different from 0.

There are many sub-classes of functions producing a sigmoidal effect. Examples:

- Pure sigmoids with smoothness characteristics expressed in trigonometric or exponential form, $f : \mathbb{R} \to [0, 1]$:

$$f(x) = \frac{1}{2}\tanh\left(\frac{\kappa x}{\pi}\right) + \frac{1}{2}$$

$$f(x) = \frac{1}{1 - e^{-ax}}$$

- Gompertz functions (a vague classification that includes above curves but can also mean special functions)
- Special functions with support in $\mathbb{R}$ such as the Error function $f : \mathbb{R} \to [0, 1]$

$$f(x) = -\frac{1}{2}\operatorname{erfc}\left(-\frac{x}{\sqrt{2}}\right)$$

- Special functions with support in $[0, 1]$, such as $f : [0, 1] \to [0, 1]$

$$f(x) = I_x(a, b),$$

  where $I_{(.)}(., .)$ is the Beta regularized function.
- Special functions with support in $[0, \infty)$

$$f(x) = Q\left(a, 0, \frac{x}{b}\right)$$

  where $Q(., ., .)$ is the gamma regularized function.

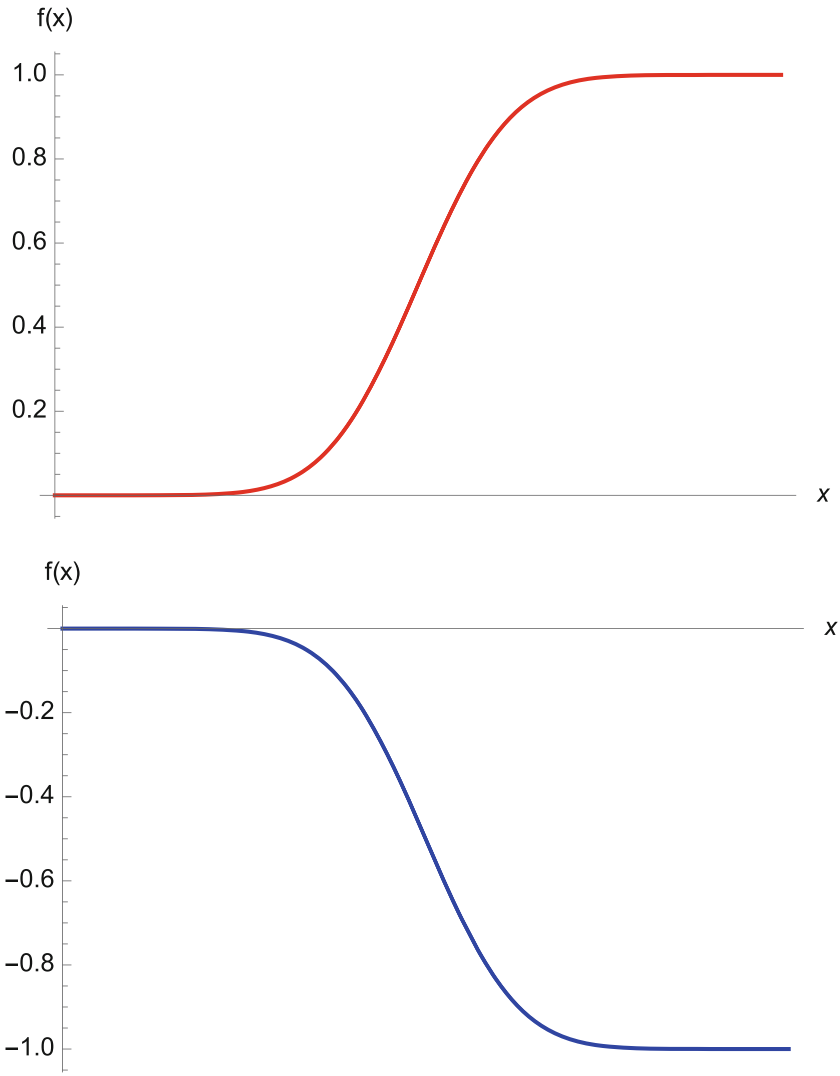


Fig. 7. Simple (first order) nonincreasing or nondecreasing sigmoids

– Piecewise sigmoids, such as the CDF of the Student Distribution

$$f(x) = \begin{cases} \frac{1}{2} I_{\frac{\alpha}{x^2+\alpha}}\left(\frac{\alpha}{2}, \frac{1}{2}\right) & x \leq 0 \\ \frac{1}{2}\left(I_{\frac{x^2}{x^2+\alpha}}\left(\frac{1}{2}, \frac{\alpha}{2}\right) + 1\right) & x > 0 \end{cases}$$

We note that the "smoothing" of the step function, or Heaviside theta $\theta(.)$ produces to a sigmoid (in a situation of a distribution or convoluted with a test

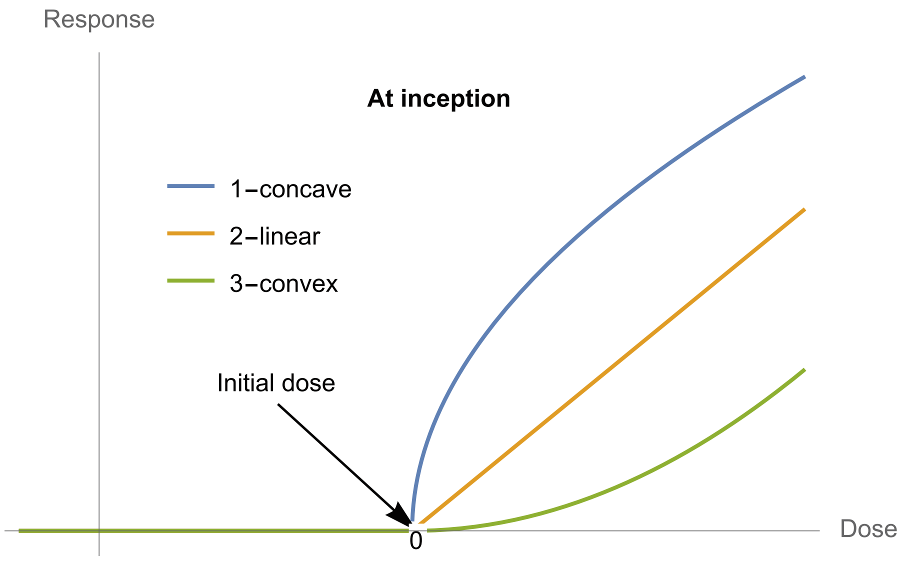


Fig. 8. The three possibilities at inception

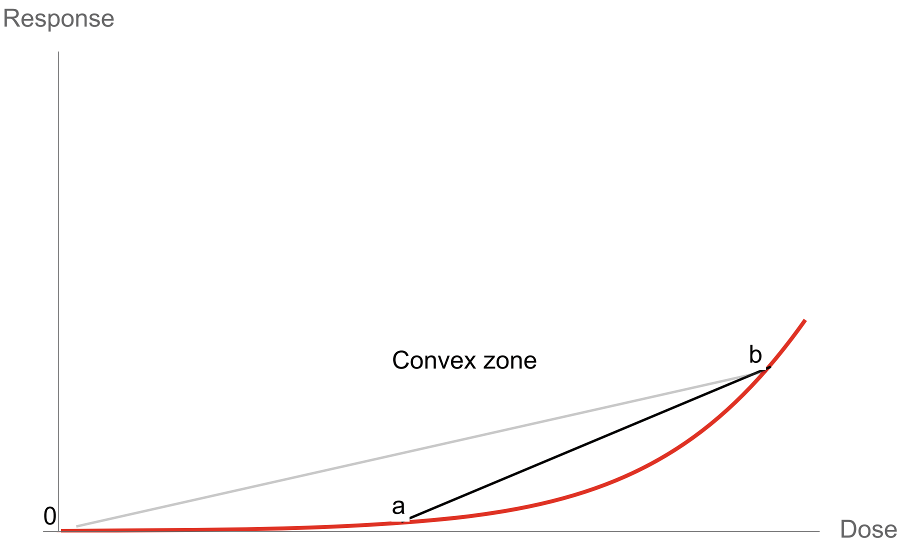


Fig. 9. Every (relatively) smooth dose-response with a floor has to be convex, hence prefers variations and concentration

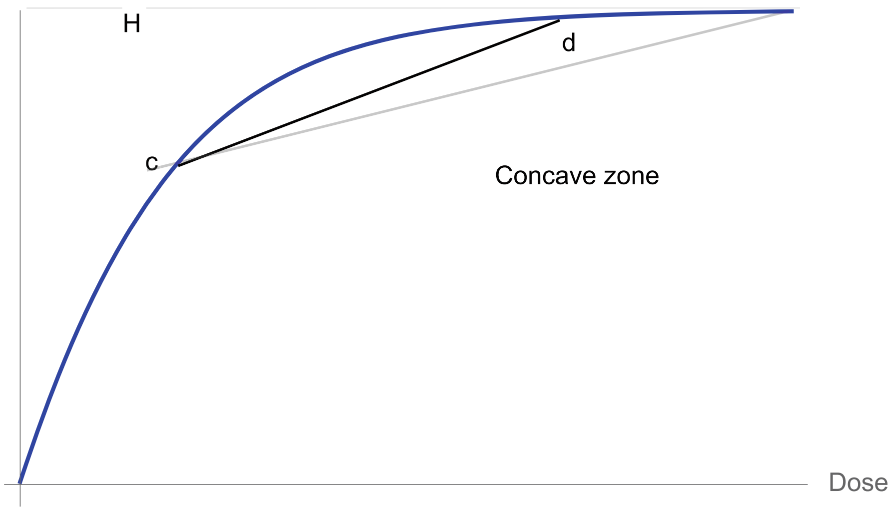


Fig. 10. Every (relatively) smooth dose-response with a ceiling has to be concave, hence prefers stability

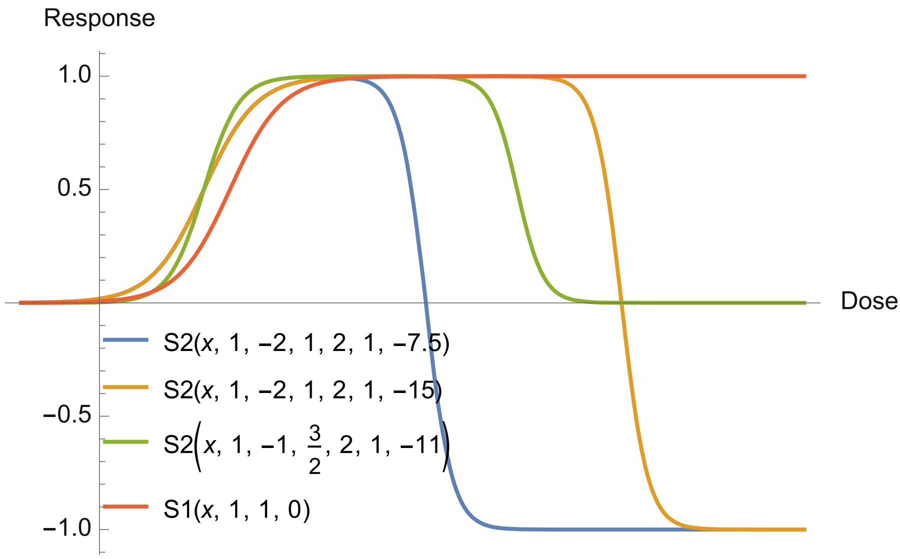


Fig. 11. The Generalized Response Curve, $S^2(x; a_1, a_2, b_1, b_2, c_1, c_2)$, $S^1(x; a_1, b_1, c_1)$ The convex part with positive first derivative has been designated as "antifragile"

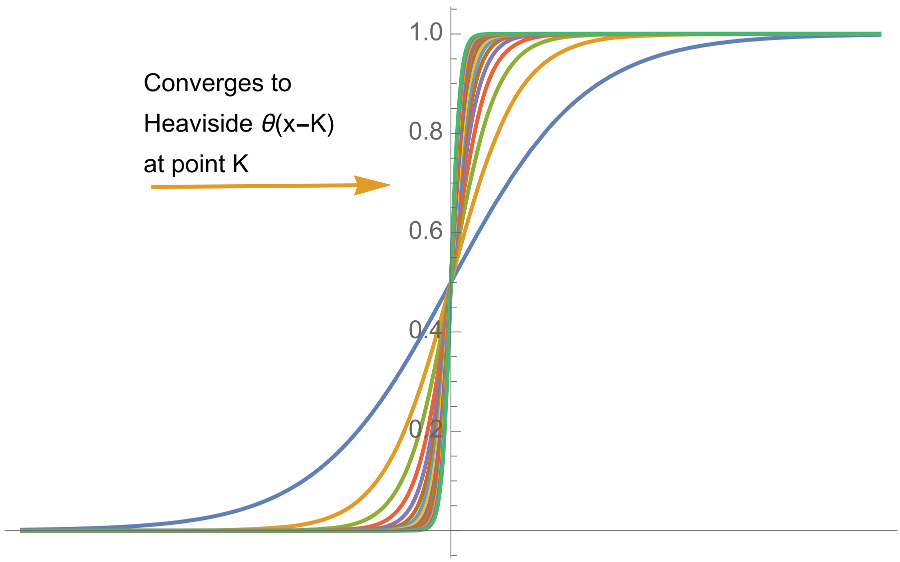


Fig. 12. The smoothing of Heaviside as distribution or Schwartz function; we can treat step functions as sigmoid so long as K, the point of the step, is different from origin or endpoint.

function with compact support), such as $\frac{1}{2}\tanh\left(\frac{\kappa x}{\pi}\right)+\frac{1}{2}$, with $\kappa \to \infty$, see Fig. 12.

1.4 Some Necessary Relations Leading to a Sigmoid Curve

Let $f_1(x) : \mathbb{R}^+ \to [0, H]$, $H \geq 0$, of class C^2 be the first order dose-response function, satisfying $f_1(0) = 0$, $f_1'(0)| = 0$, $\lim_{x\to+\infty} f_1(x) = H$, monotonic nondecreasing, that is, $f_1'(x) \geq 0 \; \forall x \in \mathbb{R}^+$, with a continuous second derivative, and analytic in the vicinity of 0. Then we conjecture that:

A - There is exist a zone $[0, b]$ in which $f_1(x)$ is convex, that is $f_1''(x) \geq 0$, with the implication that $\forall a \leq b$ a policy of variation of dosage produces beneficial effects:

$$\alpha f_1(a) + (1-\alpha) f_1(b) \geq f_1(\alpha a + (1-\alpha)b), 0 \leq \alpha \leq 1.$$

(The acute outperforms the chronic).

B - There is exist a zone $[c, H]$ in which $f_1(x)$ is concave, that is $f_1''(x) \leq 0$, with the implication that $\exists d \geq c$ a policy of stability of dosage produces beneficial effects:

$$\alpha f_1(c) + (1-\alpha) f_1(d) \leq f_1(\alpha c + (1-\alpha)d).$$

(The chronic outperforms the acute).

2 The Generalized Dose Response Curve

Let $S^N(x)$: $\mathbb{R} \to [k_L, k_R]$, $S^N \in C^\infty$, be a continuous function possessing derivatives $\left(S^N\right)^{(n)}(x)$ of all orders, expressed as an N-summed and scaled standard sigmoid functions:

$$S^N(x) \triangleq \sum_{i=1}^{N} \frac{a_k}{1 + e^{(-b_k x + c_k)}} \tag{4}$$

where a_k, b_k, c_k are scaling constants $\in \mathbb{R}$, satisfying:

(i) $S^N(-\infty) = k_L$
(ii) $S^N(+\infty) = k_R$ and (equivalently for the first and last of the following conditions)
(iii) $\frac{\partial^2 S^N}{\partial x^2} \geq 0$ for $x \in (-\infty, k_1)$, $\frac{\partial^2 S^N}{\partial x^2} < 0$ for $x \in (k_2, k_{>2})$, and $\frac{\partial^2 S^N}{\partial x^2} \geq 0$ for $x \in (k_{>2}, \infty)$, with $k_1 > k_2 \geq \ldots \geq k_N$.

By increasing N, we can approximate a continuous functions dense in a metric space, see Cybenko [8] (Figs. 9 and 10).

The shapes at different calibrations are shown in Fig. 11, in which we combined different values of N = 2 $S^2(x; a_1, a_2, b_1, b_2, c_1, c_2)$, and the standard sigmoid $S^1(x; a_1, b_1, c_1)$, with $a_1 = 1$, $b_1 = 1$ and $c_1 = 0$. As we can see, unlike the common sigmoid, the asymptotic response can be lower than the maximum, as our curves are not monotonically increasing. The sigmoid shows benefits increasing rapidly (the convex phase), then increasing at a slower and slower rate until saturation. Our more general case starts by increasing, but the response can be actually negative beyond the saturation phase, though in a convex manner. Harm slows down and becomes "flat" when something is totally broken.

3 Antifragility in the Various Literatures

Before moving to the iatrogenics section, let us review the various literature that found benefits in increase in scale (i.e. local antifragility) though without gluing their results as part of a general function.

In short the papers in this section show *indirectly* the effects of an increase in σ for diabetes, alzheimer, cancer rates, or whatever condition they studied. The scale of the distribution means increasing the variance, say instead of giving a feeding of x over each time step Δt, giving $x - \delta$ then $x + \delta$ instead, as in Eqs. 1 and 2. Simply, intermittent fasting would be having $\Delta \approx x$. And the scale can be written in such a simplified example as the dispersion $\sigma \approx \delta$.

3.1 Denial of Second Order Effect

In short, antifragility is second order effect (the average is the first order effect).

One blatant mistake in the literature lies in ignoring the second order effect when making statements from empirical data. An illustration is dietary recommendations based on composition without regard to frequency. For instance, the use of epidemiological data concerning the Cretan diet focused on composition and not how often people ate each food type. Yet frequency matters: the Greek Orthodox church has, depending on the severity of the local culture, almost two hundred vegan days per year, that is, an episodic protein deprivation; meats are eaten in lumps that compensate for the deprivation. As we will see with the literature below, there is a missing mathematical bridge between studies of *variability*, say on one hand, and the focus on food *composition* –the Longo and Fontana studies, furthermore, narrows the effect of the frequency to a given food type, namely proteins[2].

Further, the computation of the "recommended daily" units may vary markedly if one assumes second order effects: the needed average is mathematically sensitive to frequency, as we saw earlier.

3.2 Scouring the Literature for Antifragility

A sample of papers document such reaction to σ is as follows.

Mithridatization and hormesis: Kaiser [11] (see Fig. 13), Rattan [12], Calabrese and Baldwin [13–15], Aruguman et al. [16]. Note that the literature focuses on mechanisms and misses the explicit convexity argument. Is also absent the idea of divergence from, or convergence to the norm –hormesis might just be reinstatement of normalcy as we will discuss further down.

Caloric Restriction and Hormesis: Martin, Mattson et al. [17].

Treatment of Various Diseases: Longo and Mattson [18].

Cancer Treatment and Fasting: Longo et al. [19], Safdie et al. [20], Raffaghelo et al. [21], Lee et al. [22].

Aging and Intermittence: Fontana et al. [23].

For Brain Effects: Anson, Guo, et al. [24], Halagappa, Guo, et al. [25], Stranahan and Mattson [26]. The long-held belief that the brain needed glucose, not ketones, and that the brain does not go through autophagy, has been progressively replaced.

On Yeast and Longevity Under Restriction; Fabrizio et al. [27]; SIRT1, Longo et al. [28], Michan et al. [29].

For Diabetes, Remission or Reversal: Taylor [30], Lim et al. [31], Boucher et al. [32]; diabetes management by diet alone, early insights in Wilson et al. [33]. Couzin [34] gives insight that blood sugar stabilization does not have the effect

[2] Lee and Longo [9] "In the prokaryote E. coli, lack of glucose or nitrogen (comparable to protein restriction in mammals) increase resistance to high levels of H_2O_2 (15 mm) [10]".

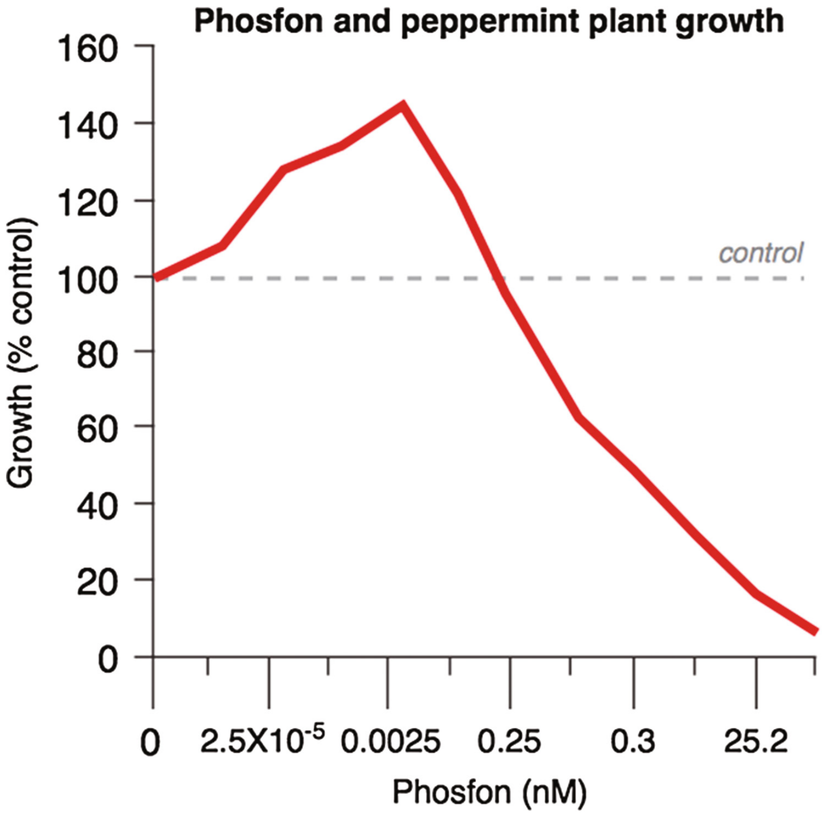


Fig. 13. Hormesis in Kaiser [11] we can detect a convex-concave sigmoidal shape that fits our generalized sigmoid in Eq. 4.

anticipated (which in our language implies that σ matters). The ACCORD study (Action to Control Cardiovascular Risk in Diabetes) found no gain from lowering blood glucose, or other metrics –indeed, it may be more opaque than a simple glucose problem remedied by pharmacological means. Synthesis, Skyler et al. [35], old methods, Westman and Vernon [36]. Bariatric (or other) surgery as an alternative approach from intermittent fasting: Pories [37], Guidone et al. [38], Rubino et al. [39].

Ramadan and Effect of Fasting: Trabelsi et al. [40]. Note that the Ramadan time window is short (12 to 17 h) and possibly fraught with overeating so conclusions need to take into account energy balance and that the considered effect is at the low-frequency part of the timescale.

Caloric Restriction: An understanding of such natural antifragility can allow us to dispense with the far more speculative approach of pharmacological interventions such as suggested in Stip (2010) –owing to more iatrogenics discussed in the next Sect. 4.

Autophagy for Cancer: Kondo et al. [41].

Autophagy (general): Danchin et al. [42], He et al. [43].

Fractional Dosage: Wu et al. [44]. Jensen's inequality in workout: Many such as Schnohr and Marott [45] compare the results of intermittent extremes with "moderate" physical activity; they got close to dealing with the fact that extreme sprinting and nothing outperforms steady exercise, but missed the convexity bias part.

Cluster of Ailments: Yaffe and Blackwell [46], Alzheimer and hyperinsulenemia as correlated, Razay and Wilcock [47]; Luchsinger, Tang, et al. [48], Luchsinger Tang et al. [49] Janson, Laedtke, et al. [50]. The clusters are of special interest as they indicate how the absence or presence of convex effect can manifest itself in multiple diseases.

Benefits of Some Type of Stress (and Convexity of the Effect): For the different results from the two types of stressors, short and chronic, "A hassle a day may keep the pathogens away: the fight-or-flight stress response and the augmentation of immune function" [51]. For the benefits of stress on boosting immunity and cancer resistance (squamous cell carcinoma), Dhabhar et al. [52], Dhabhar et al. [53], Ansbacher et al. [54].

Iatrogenics of Hygiene and Systematic Elimination of Germs: Rook [55], Rook [56] (auto-immune diseases from absence of stressor), Mégraud and Lamouliatte [57] for Helyobacter Pilori and incidence of cancer.

3.3 Extracting an Ancestral Frequency

We noted that papers such as Kaiser [11] and Calabrese and Baldwin [14], miss the point that hormesis may correspond to a "fitness dose", beyond and below which one departs from such ideal dispersion of the dose x per time period.

We can also apply the visible dose-response curve to inferring the ideal parametrization of the probability distribution for our feeding (ancestral or otherwise) and vice-versa. For instance, measuring the effects of episodic fasting on cancer, diabetes, and other ailments can lead to assessing some kind of "fitness" to an environment with a certain structure of randomness, either with the σ above or some richer measure of probability distribution. Simply, if diabetes can be controlled or reversed with occasional deprivation (a certain variance), say 24 h fasts per week, 3 days per quarter, and a full week every four years, then necessarily our system can be made to fit stochastic energy supply, with a certain frequency of deficits –and, crucially, we can extract the functional expression from such frequencies.

Note that an understanding of the precise mechanism by which intermittence works (whether dietary or in energy expenditure), which can be autophagy or some other mechanism such as insulin control, are helpful but not needed given the robustness of the mathematical link between the functional and the probabilistic.

4 Nonlinearities and Iatrogenics

Next we connect nonlinearity to iatrogenics, broadly defined as all manner of net deficit of benefits minus harm from a given intervention.

In short, Taleb and Douady [58] describes fragility as a "tail" property, that is, below a set level K, how either (1) greater uncertainty or (2) more variability translate into a degradation of the effect of the probability distribution on the expected payoff.

The probability distribution of concern has for density p, a scale s^- for the distribution below Ω a centering constant (we can call s^- a negative semideviation). To cover a broader set of distributions, we use $p_{\lambda(s)}$ where λ is a function of s.

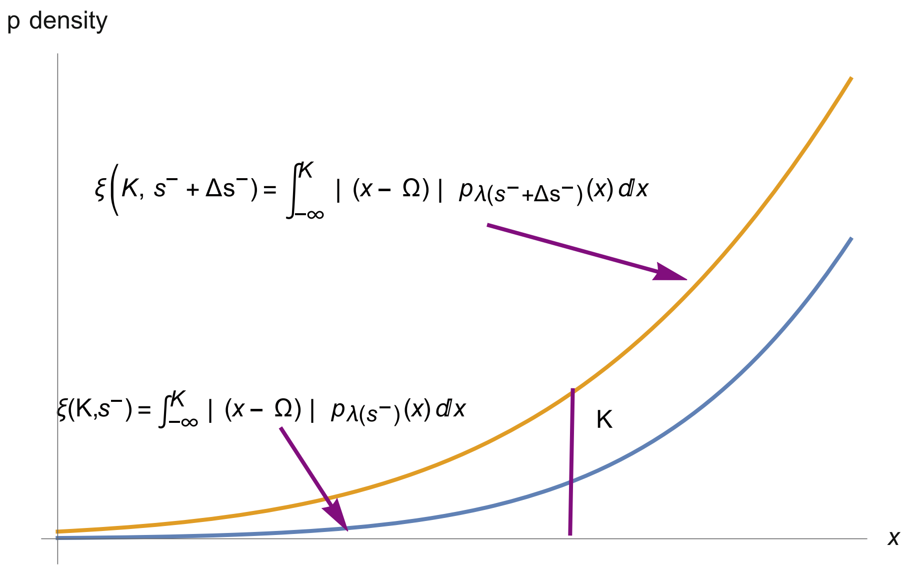


Fig. 14. A definition of fragility as left tail payoff sensitivity; the figure shows the effect of the perturbation of the lower semi-deviation s^- on the tail integral ξ of $(x - \Omega)$ below K, Ω being a centering constant. Our detection of fragility does not require the specification of p the probability distribution.

We set $\xi(.,.)$ a function of the expected value below K. Intuitively it is meant to express the harm, and, mostly its variations –one may not have a precise idea of the harm but the variations can be extracted in a more robust way.

$$\xi(s^-) = \int_{-\infty}^{K} |x - \Omega| \, p_{\lambda(s^-)(x)} \, \mathrm{d}x \tag{5}$$

$$\xi(s^- + \Delta s^-) = \int_{-\infty}^{K} |x - \Omega| \, p_{\lambda(s^- + \Delta s^-)}(x) \, \mathrm{d}x \tag{6}$$

Fragility is defined as the variations of $\xi(.)$ from an increase in the left scale s^- as shown in Fig. 14. The difference $\xi(\Delta s^-)$ represents a sensitivity to an expansion in uncertainty in the left tail.

The theorems in Taleb and Douady [58] show that:

– Convexity in a dose-response function increases ξ.
– Detecting such nonlinearity allows us to predict fragility and formulate a probabilistic decision without knowing $p(.)$.
– The mere existence of concavity in the tails implies an unseen risk.

4.1 Effect Reversal and the Sigmoid

Now let us discuss Figs. 15 and 16. The nonlinearities of dose response and hormetic or neutral effect at low doses is illustrated in the case of radiation: In Neumaier et al. [59] titled "Evidence for formation of DNA repair centers and dose-response nonlinearity in human cells", the authors write: "The standard model currently in use applies a linear scale, extrapolating cancer risk from high doses to low doses of ionizing radiation. However, our discovery of DSB clustering over such large distances casts considerable doubts on the general assumption that risk to ionizing radiation is proportional to dose, and instead provides a mechanism that could more accurately address risk dose dependency of ionizing radiation." Radiation hormesis is the idea that low-level radiation causes hormetic overreaction with protective effects. Also see Tubiana et al. [60].

Bharadwaj and Stafford present similar general-sigmoidal effects in hormonal disruptions by chemicals [61].

4.2 Nonlinearity of NNT, Overtreatment, and Decision-Making

Below are applications of convexity analysis in decision-making in dosage, shown in Figs. 15, 16 and 17.

In short, it is fallacious to translate a policy derived from acute conditions and apply it to milder ones. Mild conditions are different in treatment from an acute one.

Likewise, high risk is qualitatively different from mild risk.

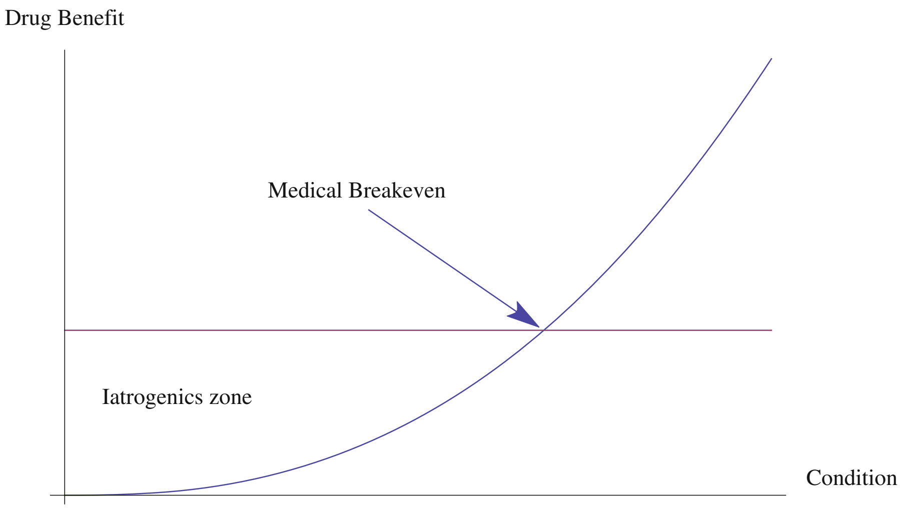


Fig. 15. Drug benefits when convex to Numbers Needed to Treat, with gross iatrogenics invariant to condition (the constant line). We are looking at the convex portion of a possibly sigmoidal benefit function.

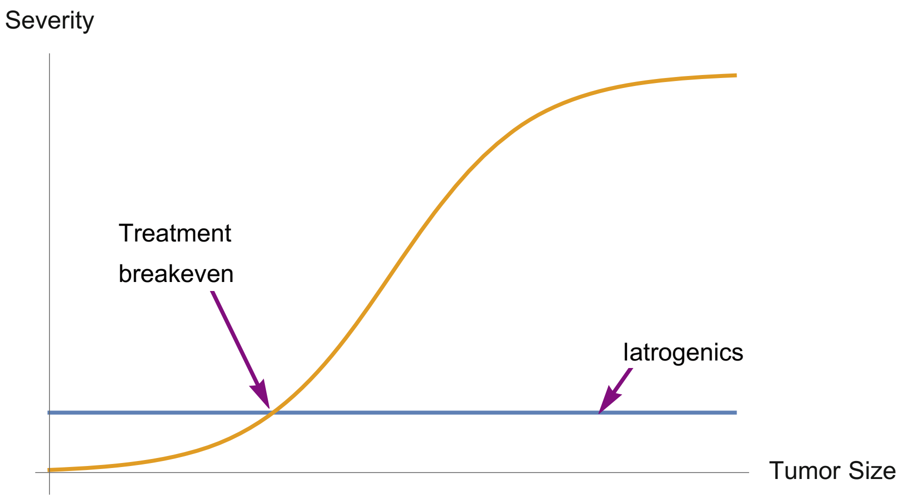


Fig. 16. Tumor breakeven we consider a wider range of Fig. 15 and apply it to the relation between tumor size and treatment breakeven.

Mammogram Controversy. There is an active literature on "overdiagnosis", see Kalager et al. [62], Morell et al. [63]. The point is that treating a tumor that doesn't kill reduces life expectancy; hence the need to balance iatrogenics and risk of cancer. An application of nonlinearity can shed some light to the

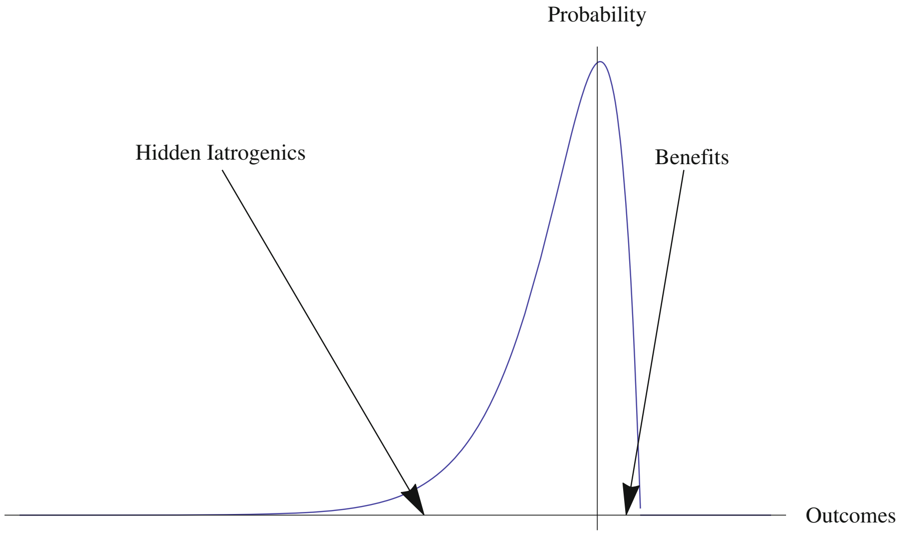


Fig. 17. Unseen risks and mild gains: translation of Fig. 15 to the skewness of a decision involving iatrogenics when the condition is mild. This also illustrates the Taleb and Douady [58] translation theorems from concavity for $S(x)$ into a probabilistic attributes.

approach, particularly that public opinion might find it "cruel" to deprive people of treatment even if it extends their life expectancy [7].

Hypertension Illustrations. Consider the following simplified case from blood pressure studies: assume that when hypertension is mild, say marginally higher than the zone accepted as normotensive, the chance of benefiting from a certain drug is close to 5% (1 in 20). But when blood pressure is considered to be in the "high" or "severe" range, the chances of benefiting would now be 26% and 72%, respectively. See Pearce et al. [64] for similar results for near-nomotensive.

But consider that (unless one has a special reason against) the iatrogenics should be safely considered constant for all categories. In the very ill condition, the benefits are large relative to iatrogenics; in the borderline one, they are small. This means that we need to focus on high-symptom conditions compare to other situations in which the patient is not very ill.

A 2012 Cochrane meta-analysis indicated that there is no evidence that treating otherwise healthy mild hypertension patients with antihypertensive therapy will reduce CV events or mortality. Makridakis and DiNicolantonio [65] found no statistical basis for current hypertension treatment. Rosansky [66] found a "silent killer" in iatrogenics, i.e. hidden risks, matching our illustration in distribution space in Fig. 17.

Statin Example. We can apply the method to statins, which appears to have benefits in the very ill segment that do not translate into milder conditions. With statin drugs routinely prescribed to lower blood lipids, although the result is statistically significant for a certain class of people, the effect is minor. "High-risk men aged 30–69 years should be advised that about 50 patients need to be treated for 5 years to prevent one [cardiovascular] event" (Abramson and Wright [67]).

For statins side effects and (more or less) hidden risks, see effects in musculoskeletal harm or just pain, Speed et al. [68]. For a general assessment, seeHilton-Jones [69], Hu, Cheung et al. [70]. Roberts [71] illustrates indirectly various aspects of convexity of benefits, which necessarily implies harm in marginal cases. Fernandez et al. [72] shows where clinical trials do not reflect myopathy risks. Blaha et al. [73] shows "increased risks for healthy patients. Also, Redberg and Katz [74]; Hamazaki et al. [75]:" The absolute effect of statins on all-cause mortality is rather small, if any."

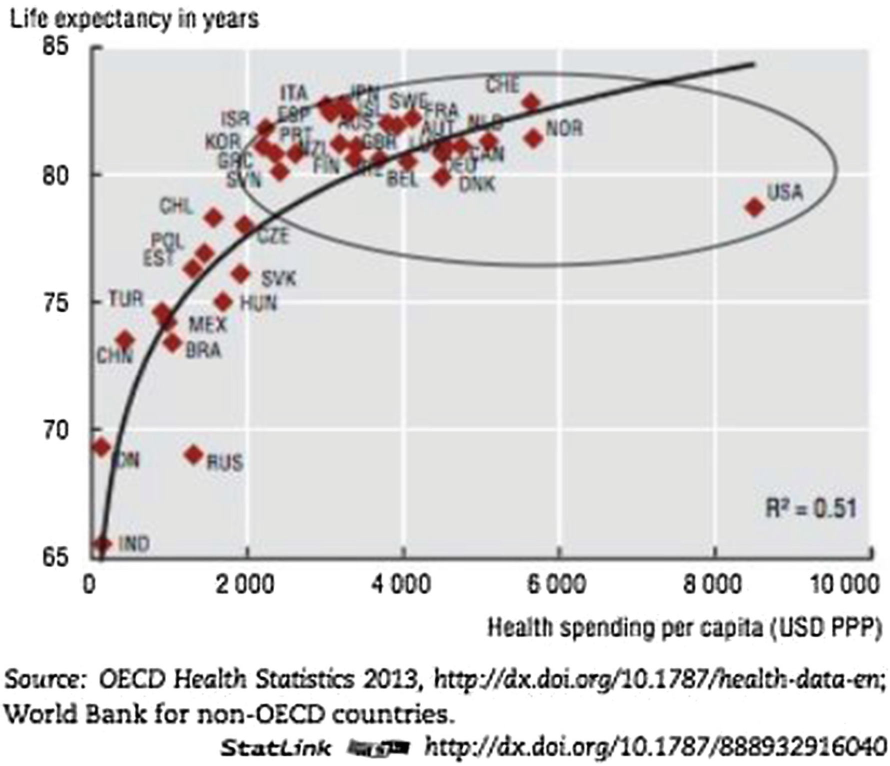


Fig. 18. Concavity of Gains to Health Spending. A more appropriate regression line than the one used by OECD should flatten off to the right, even invert to fit the USA. Credit: Edward Tufte

Other. For a similar approach to pneumonia, File [76].

Back: Overtreatment (particularly surgery) for lower back conditions is discussed in the iatrogenics (surgery or epidural), Hadler [77].

For a discussion of the application of number needed to treat in evidence-based studies, see Cook et al. [78]. One can make the issue more complicated with risk stratification (integrating the convexity to addition of risk factors), see Kannel et al. [79].

Doctor's strikes: There have been a few episodes of hospital strikes, leading to the cancellation of elective surgeries but not emergency-related services. The data are not ample ($n = 5$), but can give us insights if interpreted in *via negativa* manner as it corroborates the broader case that severity is convex to condition. It is key that there was no increase in mortality (which is more significant than a statement of decrease). See Cunningham et al. [80]. See also Siegel-Itzkovich [81]. On the other hand, Gruber and Kleiner [82] show a different effect when nurses strike. Clearly looking at macro data as in Fig. 18 shows the expected concavity: treatment results are concave to dollars invested –life expectancy empirically measured includes the results of iatrogenics.

Acknowledgment and Thanks. Harry Hong, Raphael Douady, Marco Manca, Matt Dubuque, Jacques Merab, Matthew DiPaola, Christian DiPaola, Yaneer Bar Yam, John Mafi, Michael Sagner, and Alfredo Morales.

References

1. Brewster, J.F., Graham, M.R., Mutch, W.A.C.: Convexity, Jensen's inequality and benefits of noisy mechanical ventilation. J. R. Soc. Interface **2**(4), 393–396 (2005)
2. Graham, M.R., Haberman, C.J., Brewster, J.F., Girling, L.G., McManus, B.M., Mutch, W.A.C.: Mathematical modelling to centre low tidal volumes following acute lung injury: a study with biologically variable ventilation. Respir. Res. **6**(1), 64 (2005)
3. Funk, D.J., Graham, M.R., Girling, L.G., Thliveris, J.A., McManus, B.M., Walker, E.K., Rector, E.S., Hillier, C., Scott, J.E., Mutch, W.A.C.: A comparison of biologically variable ventilation to recruitment manoeuvres in a porcine model of acute lung injury. Respir. Res. **5**(1), 22 (2004)
4. Arold, S.P., Suki, B., Alencar, A.M., Lutchen, K.R., Ingenito, E.P.: Variable ventilation induces endogenous surfactant release in normal guinea pigs. Am. J. Physiol. Lung Cell. Mol. Physiol. **285**(2), L370–L375 (2003)
5. Amato, M.B.P., Barbas, C.S.V., Medeiros, D.M., Magaldi, R.B., Schettino, G.P., Lorenzi-Filho, G., Kairalla, R.A., Deheinzelin, D., Munoz, C., Oliveira, R., et al.: Effect of a protective-ventilation strategy on mortality in the acute respiratory distress syndrome. N. Engl. J. Med. **338**(6), 347–354 (1998)
6. Jensen, J.L.W.V.: Sur les fonctions convexes et les inégalités entre les valeurs moyennes. Acta Math. **30**(1), 175–193 (1906)
7. Taleb N.N.: Antifragile: Things That Gain from Disorder. Random House and Penguin (2012)

8. Cybenko, G.: Approximation by superpositions of a sigmoidal function. Math. Control Signals Syst. **2**(4), 303–314 (1989)
9. Lee, C., Longo, V.: Fasting vs dietary restriction in cellular protection and cancer treatment: from model organisms to patients. Oncogene **30**(30), 3305–3316 (2011)
10. Jenkins, D., Schultz, J., Matin, A.: Starvation-induced cross protection against heat or H2O2 challenge in Escherichia coli. J. Bacteriol. **170**(9), 3910–3914 (1988)
11. Kaiser, J.: Sipping from a poisoned chalice. Science **302**(5644), 376 (2003)
12. Rattan, S.I.: Hormesis in aging. Ageing Res. Rev. **7**(1), 63–78 (2008)
13. Calabrese, E.J., Baldwin, L.A.: Defining hormesis. Hum. Exp. Toxicol. **21**(2), 91–97 (2002)
14. Calabrese, E.J., Baldwin, L.A.: Hormesis: the dose-response revolution. Ann. Rev. Pharmacol. Toxicol. **43**(1), 175–197 (2003)
15. Calabrese, E.J., Baldwin, L.A.: The hormetic dose-response model is more common than the threshold model in toxicology. Toxicol. Sci. **71**(2), 246–250 (2003)
16. Arumugam, T.V., Gleichmann, M., Tang, S.-C., Mattson, M.P.: Hormesis/preconditioning mechanisms, the nervous system and aging. Ageing Res. Rev. **5**(2), 165–178 (2006)
17. Martin, B., Mattson, M.P., Maudsley, S.: Caloric restriction and intermittent fasting: two potential diets for successful brain aging. Ageing Res. Rev. **5**(3), 332–353 (2006)
18. Longo, V.D., Mattson, M.P.: Fasting: molecular mechanisms and clinical applications. Cell Metab. **19**(2), 181–192 (2014)
19. Longo, V.D., Fontana, L.: Calorie restriction and cancer prevention: metabolic and molecular mechanisms. Trends Pharmacol. Sci. **31**(2), 89–98 (2010)
20. Safdie, F.M., Dorff, T., Quinn, D., Fontana, L., Wei, M., Lee, C., Cohen, P., Longo, V.D.: Fasting and cancer treatment in humans: a case series report. Aging (Albany NY) **1**(12), 988–1007 (2009)
21. Raffaghello, L., Safdie, F., Bianchi, G., Dorff, T., Fontana, L., Longo, V.D.: Fasting and differential chemotherapy protection in patients. Cell Cycle **9**(22), 4474–4476 (2010)
22. Lee, C., Raffaghello, L., Brandhorst, S., Safdie, F.M., Bianchi, G., Martin-Montalvo, A., Pistoia, V., Wei, M., Hwang, S., Merlino, A., et al.: Fasting cycles retard growth of tumors and sensitize a range of cancer cell types to chemotherapy. Sci. Transl. Med. **4**(124), 124ra27 (2012)
23. Fontana, L., Kennedy, B.K., Longo, V.D., Seals, D., Melov, S.: Medical research: treat ageing. Nature **511**(7510), 405–407 (2014)
24. Anson, R.M., Guo, Z., de Cabo, R., Iyun, T., Rios, M., Hagepanos, A., Ingram, D.K., Lane, M.A., Mattson, M.P.: Intermittent fasting dissociates beneficial effects of dietary restriction on glucose metabolism and neuronal resistance to injury from calorie intake. Proc. Natl. Acad. Sci. **100**(10), 6216–6220 (2003)
25. Halagappa, V.K.M., Guo, Z., Pearson, M., Matsuoka, Y., Cutler, R.G., LaFerla, F.M., Mattson, M.P.: Intermittent fasting and caloric restriction ameliorate age-related behavioral deficits in the triple-transgenic mouse model of alzheimer's disease. Neurobiol. Dis. **26**(1), 212–220 (2007)
26. Stranahan, A.M., Mattson, M.P.: Recruiting adaptive cellular stress responses for successful brain ageing. Nat. Rev. Neurosci. **13**(3), 209–216 (2012)

27. Fabrizio, P., Pozza, F., Pletcher, S.D., Gendron, C.M., Longo, V.D.: Regulation of longevity and stress resistance by Sch9 in yeast. Science **292**(5515), 288–290 (2001)
28. Longo, V.D., Kennedy, B.K.: Sirtuins in aging and age-related disease. Cell **126**(2), 257–268 (2006)
29. Michán, S., Li, Y., Chou, M.M.-H., Parrella, E., Ge, H., Long, J.M., Allard, J.S., Lewis, K., Miller, M., Xu, W., et al.: SIRT1 is essential for normal cognitive function and synaptic plasticity. J. Neurosci. **30**(29), 9695–9707 (2010)
30. Taylor, R.: Pathogenesis of type 2 diabetes: tracing the reverse route from cure to cause. Diabetologia **51**(10), 1781–1789 (2008)
31. Lim, E.L., Hollingsworth, K., Aribisala, B.S., Chen, M., Mathers, J., Taylor, R.: Reversal of type 2 diabetes: normalisation of beta cell function in association with decreased pancreas and liver triacylglycerol. Diabetologia **54**(10), 2506–2514 (2011)
32. Boucher, A., Lu, D., Burgess, S.C., Telemaque-Potts, S., Jensen, M.V., Mulder, H., Wang, M.-Y., Unger, R.H., Sherry, A.D., Newgard, C.B.: Biochemical mechanism of lipid-induced impairment of glucose-stimulated insulin secretion and reversal with a malate analogue. J. Biol. Chem. **279**(26), 27263–27271 (2004)
33. Wilson, E.A., Hadden, D., Merrett, J., Montgomery, D., Weaver, J.: Dietary management of maturity-onset diabetes. Br. Med. J. **280**(6228), 1367–1369 (1980)
34. Couzin, J.: Deaths in diabetes trial challenge a long-held theory. Science **319**(5865), 884–885 (2008)
35. Skyler, J.S., Bergenstal, R., Bonow, R.O., Buse, J., Deedwania, P., Gale, E.A., Howard, B.V., Kirkman, M.S., Kosiborod, M., Reaven, P., et al.: Intensive glycemic control and the prevention of cardiovascular events: implications of the accord, advance, and va diabetes trials: a position statement of the American diabetes association and a scientific statement of the American college of cardiology foundation and the American heart association. J. Am. Coll. Cardiol. **53**(3), 298–304 (2009)
36. Westman, E.C., Vernon, M.C.: Has carbohydrate-restriction been forgotten as a treatment for diabetes mellitus? A perspective on the accord study design. Nutr. Metab. **5**(1), 1 (2008)
37. Pories, W.J., Swanson, M.S., MacDonald, K.G., Long, S.B., Morris, P.G., Brown, B.M., Barakat, H.A., et al.: Who would have thought it? An operation proves to be the most effective therapy for adult-onset diabetes mellitus. Ann. Surg. **222**(3), 339 (1995)
38. Guidone, C., Manco, M., Valera-Mora, E., Iaconelli, A., Gniuli, D., Mari, A., Nanni, G., Castagneto, M., Calvani, M., Mingrone, G.: Mechanisms of recovery from type 2 diabetes after malabsorptive bariatric surgery. Diabetes **55**(7), 2025–2031 (2006)
39. Rubino, F., Forgione, A., Cummings, D.E., Vix, M., Gnuli, D., Mingrone, G., Castagneto, M., Marescaux, J.: The mechanism of diabetes control after gastrointestinal bypass surgery reveals a role of the proximal small intestine in the pathophysiology of type 2 diabetes. Ann. Surg. **244**(5), 741–749 (2006)
40. Trabelsi, K., Stannard, S.R., Maughan, R.J., Jamoussi, K., Zeghal, K.M., Hakim, A.: Effect of resistance training during ramadan on body composition, and markers of renal function, metabolism, inflammation and immunity in tunisian recreational bodybuilders. Int. J. Sport Nutr. Exer. Metab. **22**(6), 267–275 (2012)

41. Kondo, Y., Kanzawa, T., Sawaya, R., Kondo, S.: The role of autophagy in cancer development and response to therapy. Nat. Rev. Cancer **5**(9), 726–734 (2005)
42. Danchin, A., Binder, P.M., Noria, S.: Antifragility and tinkering in biology (and in business) flexibility provides an efficient epigenetic way to manage risk. Genes **2**(4), 998–1016 (2011)
43. He, C., Bassik, M.C., Moresi, V., Sun, K., Wei, Y., Zou, Z., An, Z., Loh, J., Fisher, J., Sun, Q., et al.: Exercise-induced bcl2-regulated autophagy is required for muscle glucose homeostasis. Nature **481**(7382), 511–515 (2012)
44. Wu, J.T., Peak, C.M., Leung, G.M., Lipsitch, M.: Fractional dosing of yellow fever vaccine to extend supply: a modeling study. bioRxiv, p. 053421 (2016)
45. Schnohr, P., Marott, J.L., Jensen, J.S., Jensen, G.B.: Intensity versus duration of cycling, impact on all-cause and coronary heart disease mortality: the Copenhagen city heart study. Eur. J. Cardiovasc. Prev. Rehabil. **19**, 73–80 (2011). https://doi.org/10.1177/1741826710393196
46. Yaffe, K., Blackwell, T., Kanaya, A., Davidowitz, N., Barrett-Connor, E., Krueger, K.: Diabetes, impaired fasting glucose, and development of cognitive impairment in older women. Neurology **63**(4), 658–663 (2004)
47. Razay, G., Wilcock, G.K.: Hyperinsulinaemia and alzheimer's disease. Age Ageing **23**(5), 396–399 (1994)
48. Luchsinger, J.A., Tang, M.-X., Shea, S., Mayeux, R.: Caloric intake and the risk of alzheimer disease. Arch. Neurol. **59**(8), 1258–1263 (2002)
49. Luchsinger, J.A., Tang, M.-X., Shea, S., Mayeux, R.: Hyperinsulinemia and risk of alzheimer disease. Neurology **63**(7), 1187–1192 (2004)
50. Janson, J., Laedtke, T., Parisi, J.E., O'Brien, P., Petersen, R.C., Butler, P.C.: Increased risk of type 2 diabetes in alzheimer disease. Diabetes **53**(2), 474–481 (2004)
51. Dhabhar, F.S.: A hassle a day may keep the pathogens away: the fight-or-flight stress response and the augmentation of immune function. Integr. Comp. Biol. **49**(3), 215–236 (2009)
52. Dhabhar, F.S., Saul, A.N., Daugherty, C., Holmes, T.H., Bouley, D.M., Oberyszyn, T.M.: Short-term stress enhances cellular immunity and increases early resistance to squamous cell carcinoma. Brain Behav. Immun. **24**(1), 127–137 (2010)
53. Dhabhar, F.S., Saul, A.N., Holmes, T.H., Daugherty, C., Neri, E., Tillie, J.M., Kusewitt, D., Oberyszyn, T.M.: High-anxious individuals show increased chronic stress burden, decreased protective immunity, and increased cancer progression in a mouse model of squamous cell carcinoma. PLoS ONE **7**(4), e33069 (2012)
54. Aschbacher, K., O'Donovan, A., Wolkowitz, O.M., Dhabhar, F.S., Su, Y., Epel, E.: Good stress, bad stress and oxidative stress: insights from anticipatory cortisol reactivity. Psychoneuroendocrinology **38**(9), 1698–1708 (2013)
55. Rook, G.A.: Hygiene and other early childhood influences on the subsequent function of the immune system. Dig. Dis. **29**(2), 144–153 (2011)
56. Rook, G.A.: Hygiene hypothesis and autoimmune diseases. Clin. Rev. Allergy Immunol. **42**(1), 5–15 (2012)
57. Mégraud, F., Lamouliatte, H.: Helicobacter pylori and duodenal ulcer. Dig. Dis. Sci. **37**(5), 769–772 (1992)

58. Taleb, N.N., Douady, R.: Mathematical definition, mapping, and detection of (anti)fragility. Quant. Financ. **13**, 1677–1689 (2013)
59. Neumaier, T., Swenson, J., Pham, C., Polyzos, A., Lo, A.T., Yang, P., Dyball, J., Asaithamby, A., Chen, D.J., Bissell, M.J., et al.: Evidence for formation of dna repair centers and dose-response nonlinearity in human cells. Proc. Natl. Acad. Sci. **109**(2), 443–448 (2012)
60. Tubiana, M., Aurengo, A., Averbeck, D., Masse, R.: Recent reports on the effect of low doses of ionizing radiation and its dose-effect relationship. Radiat. Environ. Biophys. **44**(4), 245–251 (2006)
61. Bharadwaj, A., Stafford III, K.C.: Hormones and endocrine-disrupting chemicals: low-dose effects and nonmonotonic dose responses. J. Med. Entomol. **47**(5), 862–867 (2010)
62. Kalager, M., Adami, H.-O., Bretthauer, M., Tamimi, R.M.: Overdiagnosis of invasive breast cancer due to mammography screening: results from the Norwegian screening program. Ann. Intern. Med. **156**(7), 491–499 (2012)
63. Morrell, S., Barratt, A., Irwig, L., Howard, K., Biesheuvel, C., Armstrong, B.: Estimates of overdiagnosis of invasive breast cancer associated with screening mammography. Cancer Causes Control **21**(2), 275–282 (2010)
64. Pearce, K.A., Furberg, C.D., Psaty, B.M., Kirk, J.: Cost-minimization and the number needed to treat in uncomplicated hypertension. Am. J. Hypertens. **11**(5), 618–629 (1998)
65. Makridakis, S., DiNicolantonio, J.J.: Hypertension: empirical evidence and implications in 2014. Open Heart **1**(1), e000048 (2014)
66. Rosansky, S.: Is hypertension overtreatment a silent epidemic? Arch. Intern. Med. **172**(22), 1769–1770 (2012)
67. Abramson, J., Wright, J.: Are lipid-lowering guidelines evidence-based? Lancet **369**(9557), 168–169 (2007)
68. Speed, W., Total, L.S.T., Care, E.B.: Statins and musculoskeletal pain (2012)
69. Hilton-Jones, D.: I-7. statins and muscle disease. Acta Myol. **28**(1), 37 (2009)
70. Hu, M., Cheung, B.M., Tomlinson, B.: Safety of statins: an update. Ther. Adv. Drug Saf. **3**, 133–144 (2012). https://doi.org/10.1177/2042098612439884
71. Roberts, B.H.: The Truth About Statins: Risks and Alternatives to Cholesterol-Lowering Drugs. Simon and Schuster, New York (2012)
72. Fernandez, G., Spatz, E.S., Jablecki, C., Phillips, P.S.: Statin myopathy: a common dilemma not reflected in clinical trials. Clevel. Clin. J. Med. **78**(6), 393–403 (2011)
73. Blaha, M.J., Nasir, K., Blumenthal, R.S.: Statin therapy for healthy men identified as "increased risk". JAMA **307**(14), 1489–1490 (2012)
74. Redberg, R.F., Katz, M.H.: Healthy men should not take statins. JAMA **307**(14), 1491–1492 (2012)
75. Hamazaki, T., Okuyama, H., Tanaka, A., Kagawa, Y., Ogushi, Y., Hama, R.: Rethinking cholesterol issues. J. Lipid Nutr. **21**(1), 67–75 (2012). https://doi.org/10.4010/jln.21.67
76. File Jr., T.M.: Another perspective: reducing the overtreatment of pneumonia. Clevel. Clin. J. Med. **80**(10), 619–620 (2013)
77. Hadler, N.M.: Stabbed in the Back: Confronting Back Pain in an Overtreated Society. University of North Carolina Press, Chapel Hill (2009)
78. Cook, R.J., Sackett, D.L.: The number needed to treat: a clinically useful measure of treatment effect. BMJ Br. Med. J. **310**(6977), 452 (1995)

79. Kannel, W.B.: Risk stratification in hypertension: new insights from the framingham study. Am. J. Hypertens. **13**(S1), 3S–10S (2000)
80. Cunningham, S.A., Mitchell, K., Narayan, K.V., Yusuf, S.: Doctors' strikes and mortality: a review. Soc. Sci. Med. **67**(11), 1784–1788 (2008)
81. Siegel-Itzkovich, J.: Doctors' strike in Israel may be good for health. BMJ **320**(7249), 1561–1561 (2000)
82. Gruber, J., Kleiner, S.A.: Do strikes kill? Evidence from New York state. National Bureau of Economic Research, Technical report (2010)

Measuring Loss of Homeostasis in Aging

Diana L. Leung[1(✉)], Linda P. Fried[2], Luigi Ferrucci[3], and Alan A. Cohen[1]

[1] Groupe de recherche PRIMUS, Department of Family Medicine, University of Sherbrooke, Sherbrooke, QC J1H 5N4, Canada
diana.leung@usherbrooke.ca
[2] Mailman School of Public Health, Columbia University, New York, NY 10032, USA
[3] Longitudinal Studies Section, National Institute on Aging, Baltimore, MD 21224, USA

Abstract. Individual biomarkers are often studied as indicators of abnormality, but a complex systems perspective suggests that further insight may be gained by considering biomarker values in the context of others. The concept of homeostasis implies that normal levels of a biomarker may be abnormal in relation to the levels of other biomarkers, and vice versa. On the premise that healthy physiological dynamics are constrained through regulation and thus converge towards certain profiles, results from our lab suggest that Mahalanobis distance (Dm), or the distance from the center of a distribution, can be used as a measure of physiological dysregulation. Specifically, Dm increases with age, and predicts mortality and many other health outcomes of age. Increase of signal with the inclusion of more biomarkers, and lack of sensitivity to biomarker choice confirm that dysregulation is indeed an emergent phenomenon. This approach can be applied at the organismal level or to specific physiological/biochemical systems. Here, in order to better understand the signal measured by Dm, we draw on the mathematical relationship between principal components (PCs) of the biomarkers and Mahalanobis distance. Our results characterize the relative distribution of biological information among the PCs, and suggest that careful removal of certain PCs in the calculation of Dm can significantly improve the biological signal.

Keywords: Dysregulation · Mahalanobis distance · Principal components

1 Introduction

The metabolic processes required to sustain life are adapted to function within certain parameters, and homeostasis can be used to refer to the ability to maintain those parameters (Meunier et al. 2014). 'Homeostasis' can also be used to refer to a physiological stable state, which, in this context, can be alternatively referred to as 'homeodynamics', which is thought to be better suited for referencing the dynamic and adaptive processes required for physiological stability (Rattan 2014). It is believed that homeodynamics have been optimized at the population level through evolutionary processes, which suggests that healthy physiological dynamics may converge towards

A. J. Morales et al. (Eds.): ICCS 2018, SPCOM, pp. 326–334, 2018.
https://doi.org/10.1007/978-3-319-96661-8_33

certain profiles (Rattan 2014). In a healthy organism, homeodynamics function on a regulatory architecture that contains considerable redundancy, which lends to its robustness (Kriete 2013). In turn, the loss of homeostasis in older age is thought to be associated with changes to physiological architecture and dynamics, the latter of which may be detected as shifted biomarker profiles from a healthy standard.

Therefore, while it is common to investigate individual biomarkers as indicators of abnormality, a systems approach using multivariate analyses can be more insightful (Cohen et al. 2018). For instance, it can detect cases where a biomarker within the normal range is abnormal in relation to other biomarker values. Mahalanobis distance (Dm) is a multivariate measure that represents the distance of a point from the center of a distribution, and our lab hypothesized that more unusual biomarker profiles might correlate with greater dysregulation (Cohen et al 2013). We found that Dm of blood biomarkers increases with age and predicts many age-related health outcomes, supporting Dm as a measure of physiological dysregulation. Furthermore, the signal of Dm increases with the number of biomarkers, but is largely insensitive to the precise choice of markers, confirming a complex systems interpretation (Cohen et al 2015a).

Here, we were interested in understanding if there were ways to improve the signal of Dm to detect physiological dysregulation. To do so, we first considered robust versions of Dm, and then tested whether reducing the information with principal components analysis (PCA) could boost the signal. PCA converts the data to an alternate coordinate system, maintaining the correlative structure within the data. Therefore, the derivation of Dm can be conceptualized in terms of either PCs or biomarkers; the Dm of all PCs is equivalent to the Dm of all biomarkers used to derive the PCs, though this connection is often overlooked (Brereton et al. 2016). We take advantage of this relationship to investigate how the signal of Dm may be improved with better characterization of its constituent PCs.

2 Materials and Methods

2.1 Data

Two datasets were used for comparison. The Women's Health and Aging Study (WHAS) is a population-based prospective study of community-dwelling women aged 65+, drawn from eastern Baltimore City and Baltimore County, Maryland (Fried et al. 2000). Invecchiare in Chianti (Aging in Chianti, "InCHIANTI") is a prospective population-based study of 1156 adults aged 65–102 and 299 aged 20–64 randomly selected from two towns in Tuscany, Italy using multistage stratified sampling in 1998. Both datasets are longitudinal with follow-up visits.

Biomarkers were chosen based on availability and with reference to previous studies (Cohen et al. 2013; Cohen et al. 2015b) to maintain consistency for comparison of results. The 36 biomarker set (those with "*" are also in the 20 biomarker set):

A/G ratio, albumin, alkaline phosphatase*, ALT*, AST*, basophil %, BUN/creatinine ratio*, calcium*, cholesterol, chloride*, creatinine*, C-reactive protein, eosinophil %, ferritin, GGT, glucose*, hemoglobin*, hematocrit*, HDL, iron, potassium*, LDH, lymphocyte %, MCH*, MCHC*, magnesium, monocyte %, neutrophil %, platelets*, red blood cell count*, RDW*, sodium*, total protein*, triglycerides, uric acid, white blood cell count**

2.2 Statistical Analyses

The biomarkers were transformed to a normal distribution if applicable, then centered and unit-scaled. PCA was performed using the *prcomp* function. Dm was calculated with the *mahalanobis* function, with the addition of the *covMcd* function from the 'robustbase' package for the robust method. Cox proportional hazards from the 'survival' package was used to model mortality. All statistical analyses were performed in R v3.4.2. Details of analytical approaches are included in the Analyses and Results section for clarity, as our subsequent methods often depended on previous results.

3 Analyses and Results

3.1 Robust Mahalanobis Distance

We compared the classical version of Dm calibrated on the full sample to a robust version using the minimum covariance determinant, which uses a specified fraction, *h*, of the population for calibration to avoid outliers and long tails. Since dysregulation is known to predict mortality (e.g. Milot et al. 2014), we assessed the biological signal of Dm through mortality hazard ratios (HRs). To see how using different fractions of the population in the estimator might affect the signal, we plotted *h* against the 95% HR (calculated as $[HR^{(Dm0.975 - Dm0.025)}]$, where Dm0.975 and Dm0.025 are the 97.5^{th} and 2.5^{th} quantiles of Dm respectively). This standardizes the analysis for the differing effect sizes of per-unit hazard ratios of the corresponding Dm measure on mortality (Fig. 1). The figure implies that classical Dm (represented by $h = 1$) outperforms robust Dm (when $h < 1$).

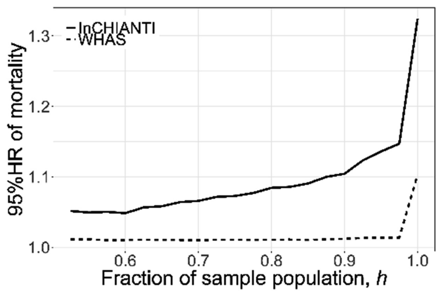


Fig. 1. HR for mortality per unit of Dm based on the fraction *h* of the population used to calculate a robust Dm for various datasets, using 36 biomarkers. Higher HR indicates stronger associations of Dm with mortality, and thus a stronger signal of dysregulation. HRs are highest in both populations for $h = 1$, indicating that signal is actually lost when using robust versions of Dm.

3.2 Characterizing PCs

Next, we attempted to understand and improve the signal of Dm by using PCs instead of the raw biomarkers. Cohen et al. (2015b) have previously used PCA as a method to examine physiological correlations among standard biomarkers. They found that the first two PCs were stably detected across demographics and were predictive of mortality and age-related health outcomes. The subsequent PCs were not further investigated due to their instability, but here we examined their biological significance in respect to variance and stability to contextualize the first PCs, and to develop a more general understanding of how PCA organizes biological information.

PC Ranking and Biological Information. To see if PC rank indicated biological significance, we calculated the effect sizes of all PC scores on mortality (Fig. 2a). While the confidence interval generally increased with increasing rank (less variance explained), 15 of 36 PCs had significant associations with mortality in WHAS, and 10 of 36 in InCHIANTI, including PCs of rank up to 33 and 27 respectively. This implies that lower-variance PCs also often hold biological information. The number of tests conducted implies substantial room for false positives in the list, but there are many more mortality-associated PCs than expected by chance.

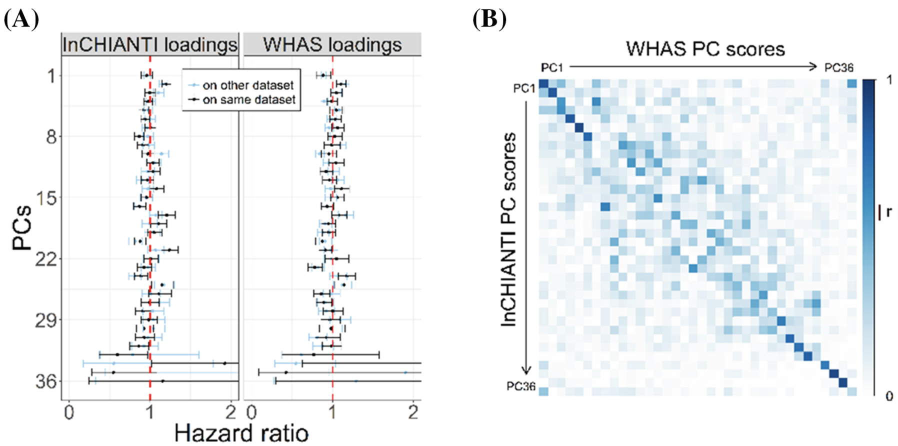


Fig. 2. (a) HRs of PCs derived from 36 biomarkers, truncated for visual clarity. Loadings derived from each dataset were applied to both datasets to obtain PC scores. (b) Correlation matrix of PC scores using WHAS and InCHIANTI PC loadings gauge stability of the PCs in the combined population.

In order to confirm the biological signal in the lower-ranked PCs, we used the loadings of these mortality-associated PCs to calculate scores for the other dataset (blue in Fig. 2a). In the lower PCs, we found that 4 PCs derived from InCHIANTI remained significant in WHAS, and 3 PCs remained significant in the reverse. Therefore, the first PCs do not seem unique in containing biological information. Also, our proxy for a biological signal - prediction of mortality - is rather crude, and many PCs unassociated with mortality may nonetheless contain important biological information.

PC Stability and Biological Information. To put the stability of the first PCs into context, and to evaluate the possibility that some of the instability may be due to shifts in ranking/order, we generated correlation matrices of the PC scores derived from analyses on varying datasets/demographics. As expected from previously published analyses, the first PCs showed relatively strong correlations with each other across datasets, and PCs became increasingly unstable with decreasing variance explained (Fig. 2b). However, it was surprising to find that the last PCs showed strong correlations across datasets as well, and that some of their correlations were even stronger than those of the first PCs. This seemed intriguing since stability, through its association with the notable first PCs, might indicate important biological processes.

3.3 Using PC Subsets to Calculate Dm

To provide an alternative gauge of biological significance in the PCs, we investigated how using different subsets of PCs in calculating Dm would affect the strength of Dm as a biological measure, again using association with mortality as a proxy. We found that omitting the last few PCs in the calculation of Dm consistently generated stronger associations with mortality than the standard of using all PCs (Fig. 3).

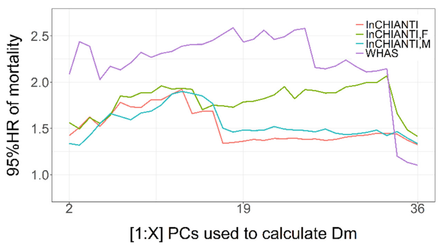


Fig. 3. Effect of using subsets of PCs to calculate Dm on HR of mortality. The x-axis represents the cumulative number of PCs, included by rank from lowest to highest, used to calculate Dm. Thus, 20 on the x-axis indicates that Dm is calculated with the first 20 PCs, etc. InCHIANTI,F and InCHIANTI,M indicate the female and male subsets, respectively.

3.4 Interpretation of the Last Few PCs

The seeming detrimental effect of the last few PCs on mortality prediction suggests that they did not contain relevant biological information - or biological information at all. This is corroborated by the examination of the correlation matrix between the PC loadings of coefficient of variance-(CV-)scaled and inversely-CV-scaled biomarkers (also known as level scaling and vast scaling, respectively), as their very low variance biomarkers did not follow as expected the inverse trend of the other PCs (van den Berg et al. 2006) (Fig. 4a). Inverse scaling would prioritize biological axes with the smallest

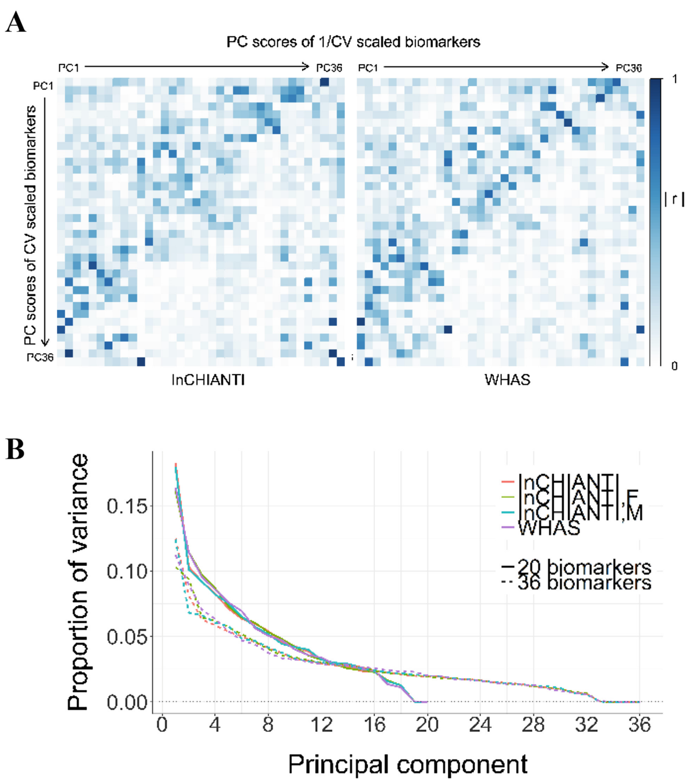


Fig. 4. (a) Correlation matrices of PC loadings derived from CV-scaled biomarkers versus PC loadings derived from inversely-CV-scaled biomarkers. CV-scaling and inverse-CV-scaling change the ranks of the PCs, such that PCs with high CVs have high or low ranks respectively. The correlations are thus expected to be strong along the negative diagonal. However, note that the last 4 PCs (33–36) are along the positive diagonal, and are thus the same axes regardless of the scaling used. (b) Scree plot of PC variance explained.

variation (i.e., highly stable), so the non-inverted trend and high correlation of the last few PCs suggest that they do not contain a biological signal.

These last few PCs also tended to correspond to a distinctive drop in variance, visually apparent on the scree plot (Fig. 4b). Others performing PCA on the InCHIANTI dataset have also observed these very-low variance PCs, and suggested that they represent measurement error, particularly as the last axis in that analysis represented the contrast between two tightly correlated receptors for TNF-α, STNRFI and STNRFII (Morrisette-Thomas et al. 2014). This interpretation agrees with the observation that the loadings of the last few PCs tend to represent the inverse relationship of highly correlated biomarkers (e.g. hemoglobin and hematocrit). These stable, very-low variance PCs might also be understood as the 'residuals' of PCA, or an effect/by-product of the correlations within the data. PCs are all linearly uncorrelated/orthogonal to each other, and PCA generates an equal number of PCs as the number of input variables, regardless of their correlation.

4 Discussion

Here we have shown that a robust version of Dm actually appears to decrease its signal in the context of detecting physiological dysregulation through biomarker profiles.

However, we have also shown how the interpretation of Dm may be improved by examining its constituent PCs, as our results suggest that: (1) biological significance is not limited to first PCs, consistent with the interpretation of Dm as an emergent phenomenon in a complex system, and (2) eliminating the last PC axes, which appear to represent measurement error or other types of noise, can boost the signal of Dm.

The interpretation of the last few PCs as noise is congruent with the relationship between the PCs and Dm. Mahalanobis distance gives all orthogonal, identically scaled variables an equal weight, with highly redundant variables downweighted. In contrast, PCA draws on the redundancies, with the highest-ranked PCs being those that contain the most redundancy (i.e., shared signal) among variables. The PCs are by definition orthogonal, and thus have equal weight when interpreting the calculation of Dm in terms of PCs. The noise of the very-low variance PCs thus carries the same per-axis weight as the information from the first few PCs, which have been previously been shown to have a clear biological interpretation (Cohen et al. 2015b). So while this 'noise' may account for a very small proportion of the data, it is greatly magnified in the Dm calculation. This would explain how its removal can have such a marked impact on the signal.

Removal of PCs is often discussed in the context of dimension reduction, where importance of PCs is mainly evaluated based on variance, like the practice of dropping PCs with variances <1% (Jolliffe 1986). In the context of using PCA in regression, it is understood that higher variance does not necessarily correspond to greater importance (Jolliffe 1986). However, discussion of PC selection for Dm appears sparse in the literature, and the literature that has been found seems to be in agreement (Brereton et al. 2016). They have also concluded that it can be advantageous to be selective of the principal components when calculating Dm. While we have observed this in the removal of the last few PCs, we have not systematically investigated other patterns of PC selection for improving Dm, and this may be a direction for future research. For example, in many -omics applications across thousands of genes, it is typical to remove the first PCs to eliminate signals due to batches, conditions, etc. that cut across multiple biological systems (Gastinel 2012, Reese et al. 2013).

This study is also limited in that it only uses associations with mortality as a proxy for biological significance. Further analyses using measures like frailty or association with age may provide greater context for interpreting these results. An additional limitation is that we do not explore the biological significance of the mid-ranked PCs. Are they simply shifting order across data sets, with measurement error and low variance explained contributing to difficulties finding strong correlations? Or is the biological information reorganized across the datasets, potentially due to factors such as population composition that might cause different processes to emerge? For example, some of our analyses suggest that it may be worth further investigating how patterns differ by sex (e.g. Fig. 3). In the end, we find it unsurprising that there is substantial biological information in the mid-ranked PCs. Biological systems are highly complex, and it would be quite surprising if a few key axes explained all the variance. However, many biological systems are contingent. For example, an acute immune response is generally mounted only in the presence of a pathogen, such that a PCA on immune parameters might generate vastly different results depending on the prevalence of pathogen infections. We should thus expect different correlation structures to emerge in organisms with differing circumstances, and we suspect that differences in population composition may account for the instability of the mid-ranked axes.

While emergent phenomena are more than the sum of their components, understanding how their components contribute to these emergent properties can provide insight to the behavior of the system. Dm seems to capture the emergent property of dysregulation, and if PCs represent physiological patterns, understanding how they impact Dm in these terms could be useful. Our current model still relies on a number of sparsely explored assumptions, such as the presence of a single optimal profile represented by the mean vector, but understanding how Dm can (and cannot) be improved should aid in our conceptualization of dysregulation, and vice versa.

References

Brereton, R.G., Lloyd, G.R.: Re-evaluating the role of the Mahalanobis distance measure. J. Chemom. **30**, 134–143 (2016)

Cohen, A.A., Legault, V., Li, Q., Fried, L.P., Ferrucci, L.: Men sustain higher dysregulation levels than women without becoming frail. J Gerontol. A **73**, 175–184 (2018)

Cohen, A.A., Li, Q., Milot, E., Leroux, M., Faucher, S., Morissette-Thomas, V., Legault, V., Fried, L.P., Ferrucci, L.: Statistical distance as a measure of physiological dysregulation is largely robust to variation in its biomarker composition. PLoS ONE **10**, e0122541 (2015a)

Cohen, A.A., Milot, E., Li, Q., Bergeron, P., Poirier, R., Dusseault-Bélanger, F., Fülöp, T., Leroux, M., Legault, V., Metter, E.J., Fried, L.P., Ferrucci, L.: Detection of a novel, integrative aging process suggests complex physiological integration. PLoS ONE **10**, e0116489 (2015b)

Cohen, A.A., Milot, E., Yong, J., Seplaki, C.L., Fülöp, T., Bandeen-Roche, K., Fried, L.P.: A novel statistical approach shows evidence for multi-system physiological dysregulation during aging. Mech. Ageing Dev. **134**, 110–117 (2013)

Gastinel, L.N.: Principal Component Analysis in the Era of « Omics » Data. Principal Component Analysis, 23 (2012)

Hubert, M., Debruyne, M.: Minimum covariance determinant. Wiley Interdisciplinary Reviews: Computational Statistics. **2**, 36–43 (2010)

Jolliffe, I.T.: Choosing a Subset of Principal Components or Variables. In: Principal Component Analysis, pp. 92–114. Springer, New York (1986)

Kriete, A.: Robustness and aging–a systems-level perspective. BioSystems **112**, 37–48 (2013)

Meunier, C.L., Malzahn, A.M., Boersma, M.: A new approach to homeostatic regulation: towards a unified view of physiological and ecological concepts. PLoS ONE **9**, e107737 (2014)

Milot, E., Morissette-Thomas, V., Li, Q., Fried, L.P., Ferrucci, L., Cohen, A.A.: Trajectories of physiological dysregulation predicts mortality and health outcomes in a consistent manner across three populations. Mech. Ageing Dev. **141–142**, 56–63 (2014)

Morrisette-Thomas, V., Cohen, A.A., Fülöp, T., Riesco, É., Legault, V., Li, Q., Milot, E., Dusseault-Bélanger, F., Ferrucci, L.: Inflamm-aging does not simply reflect increases in pro-inflammatory markers. Mech. Ageing Dev. **139**, 49–57 (2014)

Rattan, S.I.S.: Aging is not a disease: implications for intervention. Aging Dis. **5**, 196–202 (2014)

Reese, S.E., Archer, K.J., Therneau, T.M., Atkinson, E.J., Vachon, C.M., de Andrade, M., Kocher, J.-P.A., Eckel-Passow, J.E.: A new statistic for identifying batch effects in high-throughput genomic data that uses guided principal component analysis. Bioinformatics **29**, 2877–2883 (2013)

van den Berg, R.A., Hoefsloot, H.C., Westerhuis, J.A., Smilde, A.K., van der Werf, M.J.: Centering, scaling, and transformations: improving the biological information content of metabolomics data. BMC Genom. **7**, 142 (2006)

Predictive Patterns Among Microorganisms: Data Sciences for Screening Smart Bacteria for Methanogenesis and Wastewater Treatment

Charles C. Zhou[1(✉)] and Shuo Han[2]

[1] Cascade Clean Energy, Inc., Cupertino, CA 95014, USA
charles.zhou@ccleanenergy.com
[2] Chemistry Department, Missouri S&T, Rolla, MO 65409, USA
shuohan@mst.edu

Abstract. "Smart" microorganisms are named for their extraordinary ability to generate energy and materials like electricity, hydrogen, methane, cleaning wastewater and proteins. Unlocking predictive patterns between microorganisms' genetic fingerprints and their metabolism from compiled databases could provide revolutionary screening methods for discovering smart microorganisms.

In this paper, we show a self-awareness concept and theory of natural swarm intelligence (SI) that can be used to discover authoritative and popular information as well as emerging and anomalous information when traditional connections among information nodes (e.g., hyperlinks or citations) are not available. The different categories of information can be all high-value depending on the application requirements. A self-awareness of swarm intelligence is a data-driven framework, modeled and measured using a recursive distributed infrastructure of machine learning. The combination of the machine learning and swarm intelligence are extended and enhanced in a completely new perspective. We built a data model from USPTO database, NCBI database, JGI (Joint Genomic Database) and KEGG database, as well as our own bio-database.

We applied our big data biotechnology called CASCADE to microorganism populations using a measure we termed average metabolic efficiency (AME), which highly correlates with real life metabolic capabilities. We used the data models to select microbial consortia for wastewater treatment using the swarm intelligence of microbes. The collective behaviors of the selected microbes are used for cleaning wastewater and convert bio-wastes to usable energy.

In methane experiments, we found that selected microbs are not only consistent with current scientific selection, but also allowed prediction for two additional microorganisms not previously selected. This technology can potentially identify mixtures of microorganisms that work more powerfully than single ones and dramatically speed up the discovery process.

Keywords: Artificial intelligence · Big data · Swarm intelligence
Microbes · Wastewater · Anaerobic digestion

A. J. Morales et al. (Eds.): ICCS 2018, SPCOM, pp. 335–345, 2018.
https://doi.org/10.1007/978-3-319-96661-8_34

1 Introduction

Swarm intelligence (SI) is a branch of artificial intelligence, which has been existing in nature among grouping and activities of animals, birds, ants, fish or even microbes. Swarm Intelligence is the collective behavior of decentralized, self-organized systems, natural or artificial. The concept was originally used by Beni and Wang [12] in the context of cellular robotic systems.

In this paper, we show a self-awareness concept and theory of swarm intelligence [2, 8] that can be used to discover authoritative and popular information as well as emerging and anomalous information when traditional connections among information nodes (e.g., hyperlinks or citations) are not available. The different categories of information can be all high-value depending on the application requirements. A self-awareness of swarm intelligence is a data-driven framework, modeled and measured using a recursive distributed infrastructure of machine learning. The combination of the machine learning and swarm intelligence to be extended and enhanced in a completely new perspective.

Since swarm intelligence systems consist typically of a population of simple agents interacting locally with one another and with their environment. The inspiration often comes from nature, especially biological systems. The agents follow very simple rules, and although there is no centralized control structure dictating how individual agents should behave, local, and to a certain degree random, interactions between such agents lead to the emergence of "intelligent" global behavior, unknown to the individual agents [2, 8].

We used the system self-awareness and swarm intelligence to built a data model from USPTO database, NCBI database, JGI (Joint Genomic Database) and KEGG (Kyoto Encyclopedia of Genes and Genomes) database, as well as our own CASCADE (computer assisted strain construction and development engineering) database. We used these machine learning methods on the data models to select microbial consortia for wastewater treatment using the swarm intelligence of microbes. The collective behaviors of the selected microbes are used for cleaning wastewater and convert bio-wastes to usable energy.

2 Computer-Assisted Strain Construction

The swarm intelligence (SI) concept in this paper has an analogue in nature. As human being often ponder: what is the mechanism behind flocking swarms seem successfully achieve collective goals, such as looking for food or going to places in an optimized fashion even Pareto optimal using only local and simple communications. [1, 2]. Often a swarm can apply some SI to maximize a total value, e.g., get to a food target in a shortest distance. A swarm finds an optimal solution using pheromone. A pheromone is a chemical substance produced and released into the environment by a mammal or an insect which affects the behavior or physiology of others. Microbes behaves in a similar way as a form of special swarm on microscales.

In this paper, genetic information is added into the core of Computer-Assisted Strain Construction and Development Engineering (CASCADE). The first type of information, is generic gene functions or the percentage of gene usages in 23 general function categories. According to Monica Riley's classification, there are 23 gene role categories in the genome of a microorganism [1]. The number and percentage of genes for a given microorganism in every gene function category are taken from "The Comprehensive Microbial Resource (CMR)."

Second, codon usages or more granular measures for recording the amino acid composition of protein are also input with 64 codons for a given organism. Last, metabolic gene functions or the unique gene numbers in 137 metabolic function categories for a given organism are inserted. All three of these data sets are treated as input or known information to the system. More detailed functional categories can be found in the enhanced databases such as The Institute for Genetic Research (TIGR). Here one finds private experimental data can be also used for enhancement.

Figure 1 illustrates the CASCADE method which uses predictive targets such as average metabolic efficiency (AME), a measure of the average frequency that a gene appears in an organism's metabolic pathway, to compute electrogenicity, a microorganisms ability to generate clean energy like methane or electricity, where methane is a key component of biogas in anaerobic digestion process, and electricity is a key component in microbial fuel cell applications while cleaning the wastewaters.

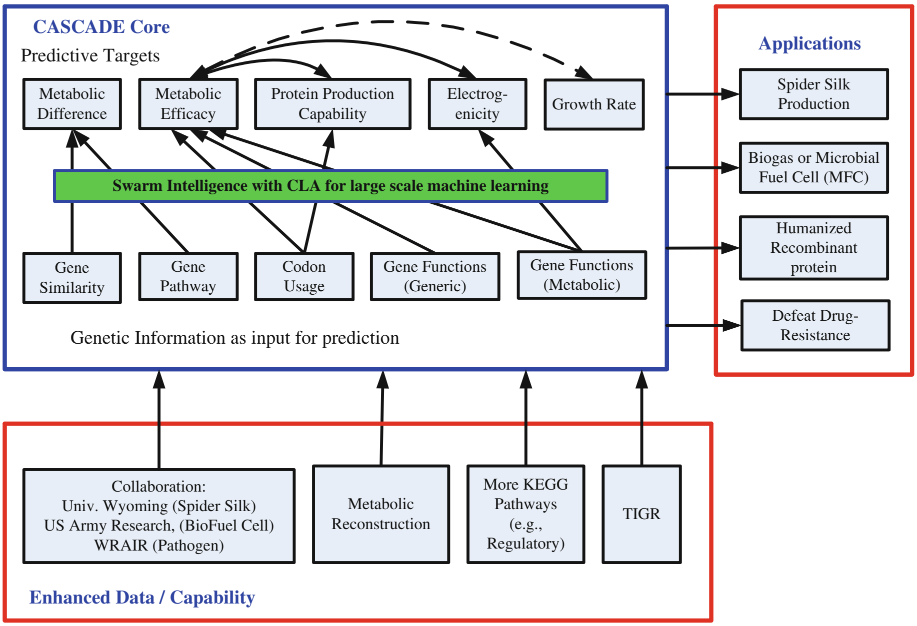


Fig. 1. Data model built with swarm intelligence and collaborative learning agents for selecting bacteria to use their collective intelligence to clean and reuse wastewater, and convert wastes to energy.

The green box in Fig. 1 shows Swarm Intelligence with collaborative learning agents (CLA) as the core modeling platform. Originally, CLA was designed and used as a large-scale machine learning system for screening efficacy and the toxicity of compounds, e.g. anti-canceanti-HIV drugs and biothreat counter measures [13]. In this paper, it is trained and validated on historical data sets (either experimental or logical) to discover knowledge patterns such as trend, recommendation, similarity and patterns to predict and recommend potentially interesting targets and properties such as methanogenicity, electrogenicity and hydrogenicity from a population of microorganisms as shown in Fig. 2. The core QIS machine learning system was wrapped around CASCADE data models constitutes the CASCADE core. The following sections detail CASCADE data models. Microbs are one of the most prolific organisms on Earth. Harnessing the power of "smart" microbs for energy generation while cleaning the wastewater and a host of other life-improving applications such as wastewater treatment has become more and more crucial to a sustainable world.

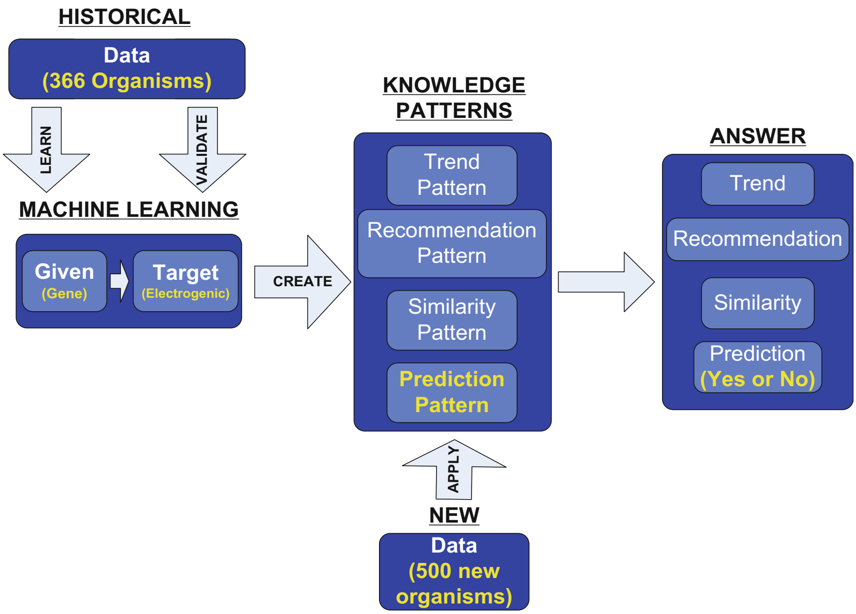


Fig. 2. Supervised leaning models of how to select microbes for our wastewater problem by using the complimentary and collective behaviors of microbes

"Smart" microbs are aptly named for their extraordinary ability to generate energy and materials like electricity, hydrogen, methane and proteins from organic sources. Enhancing how we use the unique capabilities of these microbes is important. Unlocking predictive patterns between a microorganisms genetic fingerprints and their possible "smart" metabolic capabilities opens the door to improving the interpretation of information in compiled databases of existing research which could lead to revolutionary new screening methods. For scientists, it means that microbe analysis for specific bioenergy and biotechnological uses becomes more efficient.

Machine learning, which focuses on prediction based on known properties, is the basis for the technology that we termed CASADE. This big data methodology was able to uncover predictive relationships between a microb's genetic information and its metabolic behavior.

We applied metabolic pathways from public databases, such as KEGG and investigated metabolic reconstructions for the organisms with only genomic information. Our selection included 327 bacteria in 13 groups.

We applied CASCADE technology to microb populations using a defined measure termed as average metabolic efficiency (AME), a gauge that highly correlated with metabolic capabilities that occur in real life. This measurement allowed us to explore electrogenicity for improving microbial fuel cells (MFC), methanogenicity for methane generation in anaerobic digester and protein production for spider silk.

One notable result in our work occurred with methane experiments. Here, CASCADE-selected microbs were not only consistent (5/7 overlap) with current scientific selection, but also allowed the prediction of two additional microorganisms not previously selected by conventional methods.

This machine leaning method promises to help researchers find a cocktail of mixed microorganisms that could work more efficiently and more powerfully than a single microbs. Big data research in predictive metabolomics and computational biology has the potential to speed the rate of discovery process and prediction and lower the expense of lab work and experimental trials.

3 Data Model

Average Metabolic Efficiency (AME) is the average frequency that gene of an organism is involved in a metabolic pathway, i.e. for a given organism in Table 1.

$$AME = \frac{\sum \text{Sum } (1 - \text{N})}{\sum \text{Number Unique Pathway}}$$

For example, *Acidobacteria bacterium* in Table 1 has the AME = 2.55. We found that the *AME* measure is highly correlated with metabolic capabilities in real life. It is a surrogate for representing metabolic efficiency. Higher level organisms such as a mouse (AME = 2.84) tend to have larger values of *AME*, where as lower level organisms such as microorganisms (e.g. *E. coli* AME = 2.39) tend to have smaller values of AME. A higher AME indicates a higher maintenance required for an organism for its biomass growth; therefore it might result in a higher cost to express proteins. A lower AME, on the other hand, correlates with a higher yield for expressing a gene product.

Table 1. Relations of organisms, genes to the metabolic pathways (Data shown from the 2007 KEGG database)

Organism	Enzyme (Gene)	Pathyway1	Pathyway2	Pathyway3	Pathyway4	N	Sum (1–N)	Number Unique Pathway
Aquifex aeolicus	2.3.1.157	ko00530:1				.	1	1
Aquifex aeolicus	1.3.3.4	ko00860:1				.	1	1
Aquifex aeolicus	3.1.3.2	ko00740:1	ko00361:1			.	2	2
Aquifex aeolicus	6.3.2.4	ko00550:1	ko00473:1			.	2	2
Aquifex aeolicus	2.1.3.2	ko00240:1	ko00252:1			.	2	2
.	.	.	.	.	.	.	.	.
Acideobacterial bacterium	3.5.2.9	ko00480:1				.	1	1
Acideobacterial bacterium	2.5.1.30	ko00100:1				.	1	1
Acideobacterial bacterium	3.4.13.3	ko00410:1	ko00252:1	ko00330:1	ko00340:1	.	4	4
Acideobacterial bacterium	3.3.1.1	ko00271:1	ko00450:1			.	2	2
Acideobacterial bacterium	5.4.2.1	ko00010:1				.	2	1
.	.	.	.	.	.	.	.	.
Acinetobacter sp. ADP1	4.1.2.25	ko00790:1				.	1	1
Acinetobacter sp. ADP1	3.6.3.14	ko00193:1	ko00190:1			.	2	2
Acinetobacter sp. ADP1	2.7.7.9	ko00520:1	ko00040:1	ko00500:1	ko00052:1	.	4	4
Acinetobacter sp. ADP1	3.5.1.5	ko00230:1	ko00220:1	ko00791:1		.	3	3
Acinetobacter sp. ADP1	2.5.1.6	ko00271:1	ko00450:1			.	2	2

4 Screening Microbs

Recovering methane from cleaning wastewater processing is important for two reasons: (1) methane emitted into the air after sludge digestion has an impact 23 times worse than CO_2 on the greenhouse effect (2) yet, in an opposite perspective, wastewater is an unexpected and surprisingly rich resource for electricity, hydrogen, and clean water.

In an anaerobic system [3–7] the majority of the chemical energy within a starting organic material is released as methane through the conversion of complex organic molecules to intermediate molecules. These end products also include sugars, hydrogen and acetic acid. This conversion is divided into four key biological and chemical stages, namely (i) Hydrolysis; (ii) Acidogenesis (Fermentation); (iii) Acetogenesis; (iv) Methanogenesis, and four physiologically distinct groups of microorganisms, including acetic acid-forming bacteria (acetogens) and methane-forming bacteria (methanogens), are involved in anaerobic digestion. Alternative to this scientific method, we used the CASACADE model to select smart microorganisms capable of converting sludge in wastewater to methane according to the constituents of the water.

In Fig. 3, we applied the characteristics discovered from CASCADE to query and identify microorganisms that were not previously identified as methanogenic by KEGG, literature or industries. The blue boxes highlight the microorganisms linked to the pathway characteristics (in red circles). For instance, Pyruvate/Oxoglutarate_ oxidoreductases_ High denotes higher gene activities than average are observed in the pathway Pyruvate/Oxoglutarate oxidoreductases linked to a cluster of microorganisms that are methanogens – represented in three letter abbreviations, e.g. *MAC(Methanosarcina acetivorans), MJA (Methanococcus jannaschii), MTH (Methanothermobacter thermautotrophicus)*. It is also linked to the microorganism *BJA*(*Bradyrhizobium japonicum*) in a blue box. We found three microorganisms in the population possessing both "Pyruvate/Oxoglutarate _oxidoreductases_ High" and "Tetrachloroethene_ degradation_High" which are not identified as methanogenic from the current understanding. The three microorganisms are *BJA* (*Bradyrhizobium japonicum*), *CTE*(*Chlorobium tepidum*) and *CAC* (*Clostridium acetobutylicum*). We also found *ECO* (*E. coli K-12 MG1655*) links to the group via the "ATPASE_HIGH" characteristics as shown in Fig. 3.

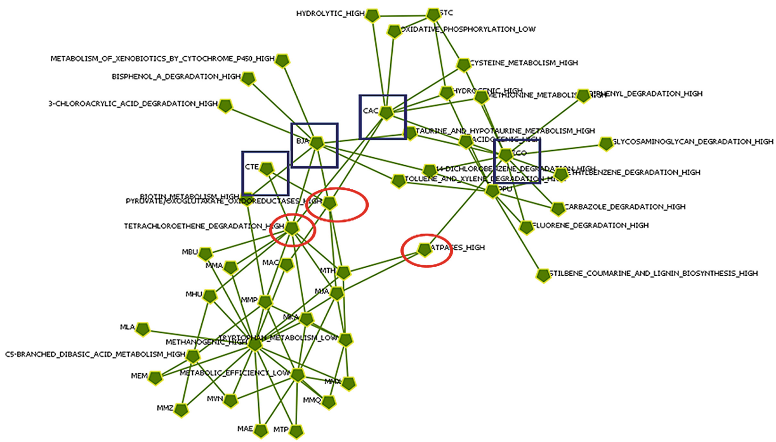


Fig. 3. Search for new microbs for methanogenic using pathway characteristics

Although the four microbs marked in Fig. 3 are not directly identified as methanogens in literature, our CASCADE model shows that they can produce methane and can function as anaerobic methanogens. It is suggested that the conversion of pyruvate to acetyl-CoA is catalyzed by pyruvate oxidoreductase in all archaebacteria. A similar pyruvate ferredoxin oxidoreductase is found in anaerobic eubacteria, therefore, it might be possible here to use a consortium to simulate anaerobic methanogens [9]. It is also suggested that the 'Pyruvate/oxoglutarate oxidoreductases' and 'C5-branched dibasic acid metabolism' are related to energy generation including valine and isoleucine biosynthesis from pyruvate (i.e. acetolactate synthase) [10]. Metabolic capacity is defined based on microbial community gene content.

After testing several anaerobic bacteria, including four strains of acetate-using methanogens, we found that *Metha-nosarcria* sp., *Methanosarcria mazei* cultures, and DCB-1 could degrade tetrachloroethene to trichloroethene. The process by which methanogens dechlorinate tetrachloroethene is a co-metabolic process and appears to be dependent on forming methane from the carbon source.

Figure 4 shows a graph similar to 4 that some substrates and products, which are critically linked to the methanogens in three letter nodes starting with ‘M’ in Fig. 4, are also linked to non-methanogens in three letter nodes in blue boxes. For example, C15489_OUT_HIGH in a red circle in Fig. 3 has four links. They represent a higher than average product C15489 (acetyl-CoA), in connection with the methanogens.

Based on this graph, we selected four microorganisms, CAC (*Clostridium acetobutylicum)*, LAC (*Lactobacillus acidophilus)*, *ECO* (*Escherichia coli K-12 MG1655*), and *PPU* (*Pseudomonas putida*) in blue boxes, which are connected with the substrates and products in red circles. These substrates and products are C15489 (acetyl-CoA), C00331 (pyruvate), C00138 (reduced ferredoxin), C00139 (oxidized ferredoxin), C02565 (N-Methylhydantoin), C00043 (UDP-N-acetylglucosamine), C00238 (potassium cation) and C00070 (copper). The four microbs could form a consortium to enhance the functions of methanogens and clean the wastewater [11].

Fig. 4. Use substrates and products to link smart microorganisms for energy generation while cleaning the wastewaters

5 Conclusions

Microorganisms are one of the most prolific organisms on Earth. Harnessing the power of “smart” microorganisms for energy generation and a host of other life-improving applications has become more and more crucial to a sustainable world.

“Smart” microorganisms are aptly named for their extraordinary ability to generate energy and materials like electricity, hydrogen, methane and proteins from organic sources. Enhancing how we use the unique capabilities of these microbes is important. Unlocking predictive patterns between a microorganisms genetic fingerprints and their possible “smart” metabolic capabilities opens the door to improving the interpretation of information in compiled databases of existing research which could lead to revolutionary new screening methods. For scientists, it means that microbe analysis for specific bioenergy and biotechnological uses becomes more efficient.

Machine learning, which focuses on prediction based on known properties, is the basis for the technology that we termed CASADE or Computer-Assisted Strain Construction and Development Engineering. This big data methodology was able to uncover predictive relationships between a microorganism’s genetic information and its metabolic behavior.

We applied metabolic pathways from public databases, such as the Kyoto Encyclopedia of Genes and Genomes (KEGG) and investigated metabolic reconstructions for the organisms with only genomic information. Our selection included 327 bacteria in 13 groups.

We applied the big data biology technology to microorganism populations using a defined measure that we termed average metabolic efficiency (AME), a gauge that highly correlated with metabolic capabilities that occur in real life. This measurement allowed us to explore electrogenicity for improving microbial fuel cells (MFC), methanogenicity for methane generation in anaerobic digester and protein production for spider silk.

One notable result in our work occurred with methane experiments. Here, CASCADE-selected microorganisms were not only consistent (5/7 overlap) with current scientific selection, but also allowed the prediction of two additional microorganisms not previously selected by conventional methods.

This big data technology promises to help researchers find a cocktail of mixed microorganisms that could work more efficiently and more powerfully than a single microorganism. Data science research in predictive metabolomics and computational biology has the potential to speed the rate of discovery process and prediction and lower the expense of lab work and experimental trials.

One challenge that arises is the large dimensionality from the potentially very large quantities of biology data–counting all possible combinations of features and attributes that might impact specific biological behavior and desired properties. Some objects and entities (e.g. number of microorganisms) available for analysis are relatively small compared to the dimensionality in research. Many of the traditional machine learning methods such as hierarchical clustering, neural networks, decision trees and association rules, are often not readily applicable. The CASCADE system draws the similarity between complex system analysis for biological data and text data, therefore, results in a more scalable and meaningful approaches.

The clustering method implemented in the CASCADE system associates elements, contexts, concepts, sequences, and clusters in a holistic manner to benefit semantic analysis. Data scoring favors identifying data anomalies that might be associated with novel biological behaviors. This is the competitive advantage of our method over other methods especially in the area of selecting smart microorganisms: A microorganism possessing a rare, novel and smart capability can be detected as data anomalies. Our method provides indicators of specific traits in a small population of microbs from a generic microb population where a large-scale of genetic information (e.g. number of genes) linking to a desired behavior like methanogenicity might emerge.

In the case of methane production and wastewater treatment, we found that biologically selected microbs and CASCADE selected microbs are consistent with five out of seven. In addition, CASCADE predicted another two microbs that are not scientifically selected. The CASCADE method can be extended to a range of applications requiring screening microorganisms that have smart capabilities in addition to methanogenicity, electrogenicity and protein productivity.

References

1. Fleischer, M.: Foundations of Swarm Intelligence: From Principles to Practice (2005). http://arxiv.org/pdf/nlin/0502003.pdf
2. Zhao, Y., Zhou, C.: System self-awareness towards deep learning and discovering high-value information. In: Proceedings of the 7th IEEE Annual Ubiquitous Computing, Electronics & Mobile Communication Conference, October 20–22, New York, USA. pp. 109–116 (2016)
3. Logan, B.E., Hamelers, B., Rozendal, R., Schröder, U., Keller, J., Freguia, S., Aelterman, P., Verstraete, W., Rabaey, K.: Microbial fuel cells: methodology and technology. Environ. Sci. Technol. **40**(17), 5181–5192 (2006)
4. Rabaey, K., Lissens, G., Siciliano, S.D., Verstraete, W.: A microbial fuel cell capable of converting glucose to electricity at high rate and efficiency. Biotechnol. Lett. **25**, 1531–1535 (2003)
5. Niessen, J., Schröder, U., Rosenbaum, M., Scholz, F.: Fluorinated polyanilines as superior materials for electrocatalytic 8 anodes in bacterial batteries. Electrochem. Commun. **6**, 571–575 (2004a)
6. Niessen, J., Schröder, U., Scholz, F.: Exploiting complexcarbohydrates for microbial electricity generation – a bacterial fuel cell operating on starch, Electrochem. Commun. **6**, 955–958 (2004b)
7. Kim, M.J., Cho, H.S., Kim, J.Y.: Anaerobic biodegradability of plastic garbage bags based on starch polymer and aliphatic polyester. In: Proceedings of Sardinia 2007, Eleventh International Waste Management and Landfill Symposium, S. Margherita di Pula, Cagliari, Italy, 1–5 October 2007, pp. 517–518 (2007)
8. Zhou, C., Zhao Y.: Method and system for knowledge pattern search and analysis for selecting microorganisms based on desired metabolic property or biological behavior. US patents 9,026,373 and 9,792,404
9. Kates, M., Kushner, D.J., Matheson, A.T. (eds.): The Biochemistry of Archaea (Archaebacteria). Elsevier, Amsterdam (1993)

10. Turnbaugh, P.J., Ley, R.E., Mahowald, M.A., Magrini, V., Mardis, E.R., Gordon, J.I.: An obesity-associated gut microbiome with increased capacity for energy harvest. Nature **444**, 1027–1031 (2006)
11. Zhou, C., Killgrow, S., Hill, C.: Cascade Clean Energy System for Water & Wastewater Treatment. www.energy.ca.gov/2015publications/CEC-500-2015-065.pdf
12. Beni, G., Wang, J.: Swarm intelligence in cellular robotic systems, proceed. In: NATO Advanced Workshop on Robots and Biological Systems, Tuscany, Italy, June 26–30, 1989
13. Zhao, Y., Wei, S., Oglesby, I., Zhou, C.: Utilizing the quantum intelligence system for drug discovery (QIS D2) for Anti-HIV and Anti-Cancer cocktail detection. J. Med. Chem. Bio. Res. Def., 7 (2009)

The Effect of Removal of Self-loop for Attractor in Cell Cycle Network

Shu-ichi Kinoshita[1,2(✉)] and Hiroaki Yamada[3]

[1] Department of Mathematical Engineering, Faculty of Engeneering at Musashino University, 3-3-3 Ariake Koutou-ku, Tokyo 135-0063, Japan
kino@musashino-u.ac.jp
[2] Institute for Advanced Study of Mathematical Sciences (MIMS), Meiji University, 4-21-1 Nakano, Nakano-ku, Tokyo 164-8525, Japan
[3] Yamada Physics Research Labolatory (YPRL), 5-7-14-205 Aoyama, Niigata 950-2002, Japan
hyamada@uranus.dti.ne.jp

Abstract. We numerically investigated the role of degenerate self-loops on the attractors and its basin size of the ES cell network of *C.elegans*. It is found that for the point attractors a simple division rule of the state space is caused by removing the self-loops of the network.

Keywords: Gene regulatory network · Attractors · ES cell network *C.elegans*, Degenerate self-loop

1 Introduction

Recently, some networks representing metabolic reactions in the cell and gene regulatory responses through transcription factors have been elucidated along with progress of experimental systems and accumulation technology. Kauffman *et al.* modeled the early cells before differentiation with the dynamics of the network, and made the type of the attractors correspond to the type of cells after the differentiation [1–3] On the other hand, Li *et al.* discovered that in the model of the gene regulatory network related to the cell-cycle, there is a fixed point with a very large basin size, and the transition process to the fixed point corresponds to the expression pattern of the gene in each process of the cell-cycle [4] In this report, using the ES cell network of the *C.elegans* [5], we investigate the relationship between the fixed points (point attractors) with large basin size and the presence of the self-loops in the network. It should be noticed that in the network of the Kauffman *et al.*, there is no self-regulating factor (self-loop), but in the model of Li *et al.* the existence of the self-loops has a great influence on the attractors.

2 Model

Let us take the binary value $\{0, 1\}$ as the state S_i of each node i corresponding to the numbered genes, $i = 1, 2, ..., N$, where N is the total number of the nodes.

A. J. Morales et al. (Eds.): ICCS 2018, SPCOM, pp. 346–351, 2018.
https://doi.org/10.1007/978-3-319-96661-8_35

The states 1 and 0 correspond to expressed and unexpressed genes, respectively, and the attractors of the dynamics are associated to cell differentiation. The effect on the node i from the node j is defined as $B_i = \sum_j^N a_{ij} S_j$ and a_{ij} denotes matrix element of the weighted adjacency matrix A representing the interaction between the genes. We take $a_{ij} = +1$ when the node j positively regulates the node i (positive interaction), and $a_{ij} = -1$ when the node j negatively suppresses the node i (negative interaction).

The states of node follow a threshold dynamics from discrete time t to $t+1$ ($t \in \mathbf{N}$) by using the parallel updating scheme:

$$S_i(t+1) = \begin{cases} 0 & (B_i(t) < \theta_i) \\ 1 & (B_i(t) > \theta_i) \\ S_i(t) & (B_i(t) = \theta_i), \end{cases} \tag{1}$$

where θ_i denotes the threshold value of the node i. The ES cell network of the *C.elegans* (denoted by $G^{(0)}$) by Huang *et al.* is a special one in a sense that all nodes of the existing self-loops are given as $a_{ii} = -1$. The network is shown in Fig. 1. We take the values $\theta_i = 0$ for all i in this report. Each regulatory

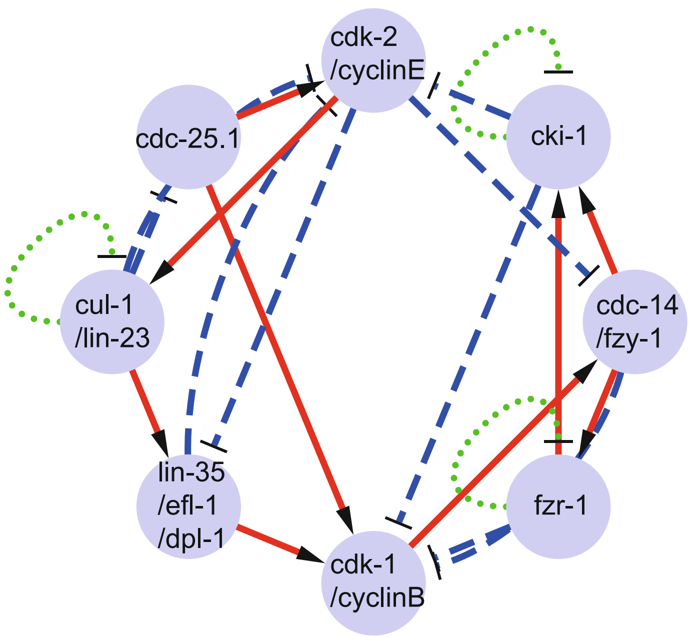


Fig. 1. Gene regulatory network $G^{(0)}$ of the ES cell of *C.elegans* [5]. Each circle represents a protein (cycling or transcription factor) involved in the gene regulation. For the links connecting the respective proteins, the blue-dashed lines represent the effect of the suppression control, and the red-solid lines represent the effect of the activation control. In addition, the self-loops by green-dotted lines represent the effect of protein degradation in the absence of external input.

factor is represented by each numbered node ($i = 1, 2, ..., 8$). There are self-degeneration loops on the 3 nodes, *cki-1*, *fzr-1*, *cul-1/lin-23*. In this network, the total state number is $W = 2^8 = 256$, and all steady states are 5 point attractors by numbering as $\mathbf{A}^{(0)} = \{A_1^{(0)}, A_2^{(0)}, A_3^{(0)}, A_4^{(0)}, A_5^{(0)}\}$, as shown in Table 1. The state of the point attractor with the largest basin size among these is $A_1^{(0)} = 000010111 = 23$, where the last number is in decimal. Figure 2 shows the basin structure of the 256 initial states flowing to the fixed points given in the Table 1. The red circles are the point attractors of $G^{(0)}$.

Table 1. 5 attractors in the original gene regulatory network. of the *C.elegans* (All are point attractors.) The second line shows that there is a degenerate self-loop when mark @ is present in the node. In the decimal notation, each attractor is displayed as, $A_1^{(0)} = 23$, $A_2^{(0)} = 87$, $A_3^{(0)} = 7$, $A_4^{(0)} = 71$, $A_5^{(0)} = 0$, The last column (BS) represents the basin size of the attractors.

No	1	2	3	4	5	6	7	8	BS
			@			@		@	
	cdk-2/ cyclinE	cdc-25.1	cul-1/ lin-23	lin-35/ efl-1/dpl-1	cdk-1/ cyclinB	frz-1	cdc-14/ fzy-1	cdi-1	
$A_1^{(0)}$	0	0	0	1	0	1	1	1	219
$A_2^{(0)}$	0	1	0	1	0	1	1	1	16
$A_3^{(0)}$	0	0	0	0	0	1	1	1	12
$A_4^{(0)}$	0	1	0	0	0	1	1	1	5
$A_5^{(0)}$	0	0	0	0	0	0	0	0	4

3 Numerical Result

In this section, we investigate the effect of the degenerate self-loops on the attractors of the original network $G^{(0)}$. Therefore, we write the network from which the degenerate self-loop of the kth node is removed from $G^{(0)}$ as $G^{(-k)}$, k selects from the nodes with the self-loop. The attractor sets are indicated as $\mathbf{A}^{(-k)} = \{A_1^{(-k)}, A_2^{(-k)}, ..., ..A_{n_{-k}}^{(-k)}\}$, where n_{-k} means the number of attractors in the networks $G^{(-k)}$.

We show in the Table 2 all point attractors of the gene regulatory network $G^{(-8)}$ which removed the degenerate self-loop of cki-1 (the 8th node). Figure 3 shows the basin structure of $G^{(-8)}$. We compare the attractors in the network $G^{(-8)}$ with those in $G^{(0)}$. It is found that $A_1^{(-8)} = A_1^{(0)}$, $A_3^{(-8)} = A_2^{(0)}$, $A_4^{(-8)} = A_3^{(0)}$, $A_5^{(-8)} = A_4^{(0)}$, $A_8^{(-8)} = A_5^{(0)}$. That is, all of the attractor sets $\mathbf{A}^{(0)}$ of the original network $G^{(0)}$ is included the attractor set $\mathbf{A}^{(-8)}$ of the network $G^{(-8)}$, i.e. $\mathbf{A}^{(0)} \subset \mathbf{A}^{(-8)}$.

Next, we focus on the change of the basin size. It follows that the basin size of the attractor $A_1^{(0)}$ with the largest basin size is reduced by the elimination of the degenerate self-loop. Also, the basin size of the other attractors are also reduced from those of $\mathbf{A^{(0)}}$. In Fig. 2 the green squares indicate the three point attractors newly added by the network becoming $G^{(-8)}$. Obviously, the basin size of the same attractor of $G^{(-8)}$ to those of attractor of $G^{(0)}$ is smaller than those of $G^{(0)}$, and they are caused by branching from the basin of $G^{(0)}$. Accordingly, it is also easy to understand that all attractors (attractor sets) of the original network $G^{(0)}$ are included in the attractor set of $G^{(-8)}$.

Although above results are for the specific case which the degenerate self-loop of the gene cki-1 has been removed, but also it is found that the similar results are also true for the cases removing the other degenerate self-loops, i.e. $\mathbf{A}^{(0)} \subset \mathbf{A}^{(-k)}$, when the self-loop is removed from the $k-$th node. Furthermore, if we apply this rule repeatedly in the process of removing the self-loops, we can see that in general the above relations of the attractors and the basin size stands for the relationship before and after removing the self-loops.

Table 2. Attractors in the gene regulatory network $G^{(-8)}$ which removed the degenerate self-loop of cdi-1(the 8th node). (All are point attractors.) The last column (BS) represents the basin size of the attractors. In the decimal notation, each attractor is displayed as, $A_1^{(-8)} = 23$, $A_2^{(-8)} = 17$, $A_3^{(-8)} = 87$, $A_4^{(-8)} = 7$, $A_5^{(-8)} = 71$, $A_6^{(-8)} = 1$, $A_7^{(-8)} = 65$, $A_8^{(-8)} = 0$.

No	1	2	3 @	4	5	6 @	7	8	BS
$A_1^{(-8)}$	0	0	0	1	0	1	1	1	129
$A_2^{(-8)}$	0	0	0	1	0	0	0	1	87
$A_3^{(-8)}$	0	1	0	1	0	1	1	1	16
$A_4^{(-8)}$	0	0	0	0	0	1	1	1	12
$A_5^{(-8)}$	0	1	0	0	0	1	1	1	6
$A_6^{(-8)}$	0	0	0	0	0	0	0	1	3
$A_7^{(-8)}$	0	1	0	0	0	0	0	1	2
$A_8^{(-8)}$	0	0	0	0	0	0	0	0	1

4 Summary and Discussion

In this short report, we investigated the influence of the degenerate self-loop on attractor of the gene regulatory network model of the ES cell network of the *C.elegans*, in order to clarify the relationship between the point attractor with a large basin and the existence of the degenerate self-loops. In the case of networks with degenerate self-loops removed from the original network $G^{(0)}$, only the point

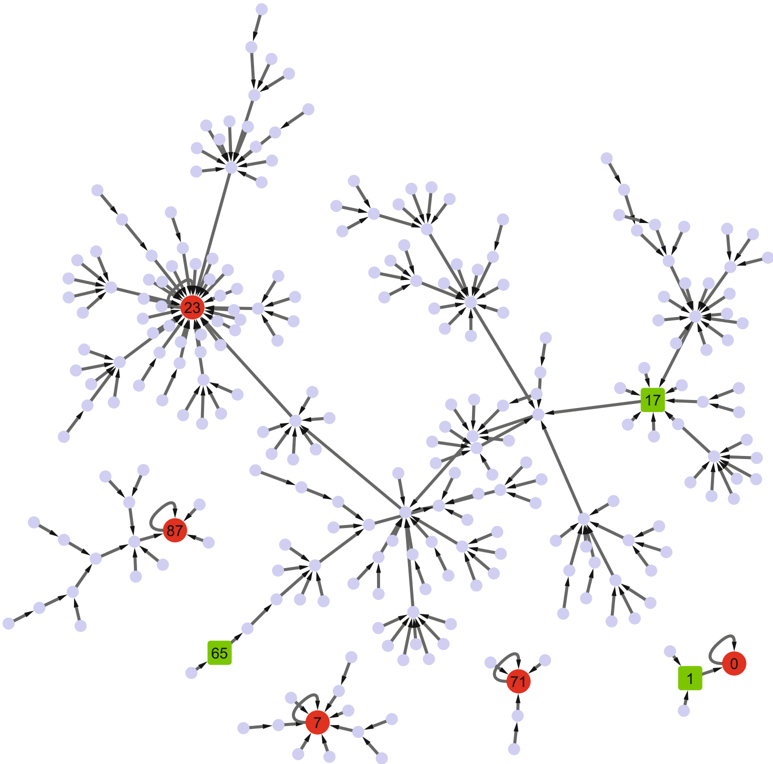


Fig. 2. The point attractors and the basin structures of the network $G^{(0)}$. The 5 red circles present the common point attractors to $G^{(0)}$ and $G^{(-8)}$. The green squares present attractors newly added when the network becoms $G^{(-8)}$. Each number in a circle or square is the decimal number of the attractor.

attractor appears because all of the self-loops are degenerate. The attractor set of the network without the degenerate self-loops includes all attractors of the original network $G^{(0)}$.

Above result was consistent with those in the gene regulatory network of the cell-cycle of budding yeast that the attractors are only fixed points [6]. On the other hand, it should be noticed that in general, in the gene regulatory networks with both the self-regression loops and self-activation loops, limit cycles with the period more than 2 may appear, in addition to the point attractors. For a discussion of these cases, see the Ref. [6].

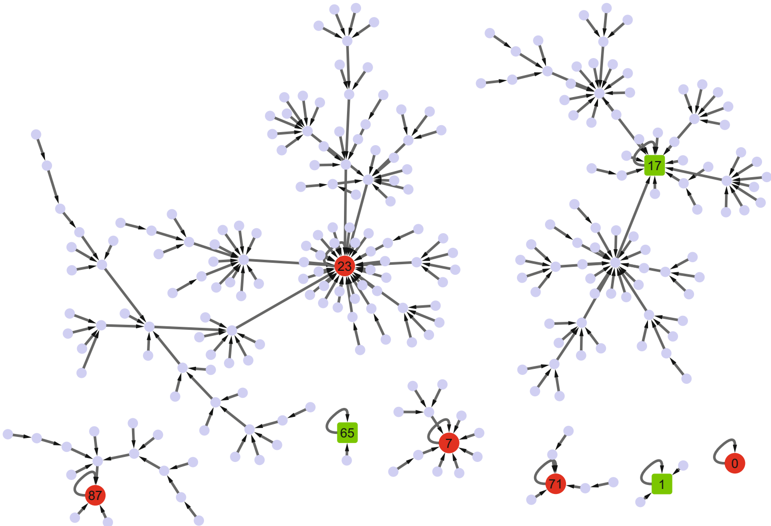


Fig. 3. The point attractors and the basin structures of the network $G^{(-8)}$. Each number in a circle or square is the decimal number of the attractor.

References

1. Kauffman, S.A.: Metabolic stability and epigenesis in randomly constructed genetic nets. J. Theor. Biol. **22**, 437–67 (1969)
2. Iguchi, K., Kinoshita, S., Yamada, H.S.: Boolean dynamics of Kauffman model with a scale-free network. J. Theor. Biol. **247**, 138–151 (2007)
3. Kinoshita, S., Iguchi, K., Yamada, H.S.: Intrinsic properties of Boolean dynamics in complex networks. J. Theor. Biol. **256**, 351–369 (2009)
4. Li, F., et al.: The yeast cell-cycle network is robustly designed. Proc. Natl. Acad. Sci. USA **101**, 4781–4786 (2004)
5. Huang, X., et al.: Boolean genetic network model for the control of C. elegans early embryonic cell cycles, BioMed. Eng. OnLine 12(Suppl 1) S1 (2013)
6. Kinoshita, S., Yamada, H.: Role of self-loop in cell-cycle network of budding yeast (preprint)

Network Modularity and Hierarchical Structure in Breast Cancer Molecular Subtypes

Sergio Antonio Alcalá-Corona[1,2,3(✉)], Guillermo de Anda-Jáuregui[1], Jesús Espinal-Enriquez[1,2], Hugo Tovar[1], and Enrique Hernández-Lemus[1,2]

[1] National Insitute of Genomic Medicine, Mexico City, Mexico
[2] Centro de Ciencias de la Complejidad (C3), Universidad Nacional Autónoma de México, Ciudad de México, Mexico
[3] Department of Ecology & Evolution, The University of Chicago, Erman Biology Center, 1103 East 57th Street, Chicago, IL 6063, USA
salcala@uchicago.edu

Abstract. Breast Cancer is the malignant neoplasm with the highest incidence and mortality among women worldwide. It is a heterogeneous and complex disease, its classification in different molecular subtypes is a clear manifestation of this. The recent abundance of genomic data on cancer, make possible to propose theoretical approaches to model the process of genetic regulation. One of these approaches is gene transcriptional networks which represent the regulation and co-expression of genes as well-defined mathematical objects. These complex networks have global topological and dynamic properties. One of these properties is modular structure, which may be related to known or annotated biological processes. In this way, different modular structures in transcription networks can be seen as manifestations of regulatory structures that closely control some biological processes. In this work, we identify modular structures on gene transcriptional networks previously inferred from microarray data of molecular subtypes of breast cancer: luminal A, luminal B, basal, and HER2-enriched. Using a methodology based on the identification of functional modules in transcriptional networks, we analyzed the modules (communities) found in each network to identify particular biological functions (described in the Gene Ontology database) associated to them. We also explored the hierarchical structure of these modules and their functions to identify unique and common characteristics that could allow a better level of description of such molecular subtypes of breast cancer. This approach and its findings are leading us to a better understanding of the molecular cancer subtypes and even contribute to direct experiments and design strategies for their treatment.

Keywords: Breast cancer subtypes · Modular structure
Gene regulatory networks

S. A. Alcalá-Corona, G. de Anda-Jáuregui, J. Espinal-Enriquez and E. Hernández-Lemus—Equal Contributors.

A. J. Morales et al. (Eds.): ICCS 2018, SPCOM, pp. 352–358, 2018.
https://doi.org/10.1007/978-3-319-96661-8_36

1 Modularity in Biological Networks

Biological systems are comprised of different types of molecules that interact in a variety of fashions. These biological systems may be modelled as *biological networks*. These may be defined as a graph $(G(V,E))$ over a duplex formed by two sets: a set of nodes or vertices (V) representing *biomolecules*, and a set of edges or links (E) that represent physical or chemical interactions among such biomolecules [5]. Studying biological systems from a network perspective allows the use of well-developed mathematical tools, in such a way that topological properties of the network model may be associated to features of the biological system.

Modularity is a property that has been widely observed in biological networks [15]. Biological networks often exhibit non-random connectivity patterns. Biological networks also exhibit heterogeneity in terms of degree distribution: in other words, there is a notable difference in the number of nodes with many and few neighbors. Furthermore, these networks also exhibit high heterogeneity in the amount of links at a global and local scale, with regions that have a high edge density alongside regions of low edge density. These patterns of organization lead to the emergence of structural subunits known as *modules* or *communities*. Beyond a formal definition, the concept of a module inside a network can describe a subgraph which contains a higher number of edges between the nodes that belong to the module, than the number of links to nodes outside the module.

The concept of modularity in biological systems agrees with the fact that processes necessary for the survival and development of life are often organized and compartmentalized through either location, functionality, or a combination of both. With this in mind, it is of current scientific interest to explore to what extent modules identified in biological networks may reflect these functional organization.

Many methodologies have been developed in order to reconstruct biological networks [11,14,16] and study modularity in them. We have focused in the last years in using an approach that relies on Information Theory-based methods for network deconvolution and module detection in the context of transcriptional networks [9]. Transcriptional networks attempt to reconstruct the gene *regulatory program* associated to specific cellular phenotypes using gene expression data (from high-throughput technologies such as microarrays or RNA-seq).

Mutual Information (MI) may be used as a measure of statistical dependence between the expression levels of two genes (or more generally, any two measured transcripts) in order to construct a network: the underlying hypothesis is that high levels of MI are indicative of co-regulation of a gene pair. With this in mind, the emergence of modules in such a network could indicate a common regulation of said group of genes, which may in turn be related to a functional aspect associated to the phenotype. While the problem of identifying modules in large networks remains open, and many methods have been developed (see [7,8]), Information Theory-based methods for module detection are available; our group has often used an implementation of the Infomap algorithm [12], which has been demonstrated to obtain an almost optimal network partition.

We will now present previously published [2,3] examples of the study of breast cancer biology through the exploration of modularity in transcriptional networks. Particularly, we will show how network features can be used to identify differences between clinical manifestations of the disease, as embodied in the molecular subtype classification [10], and how some of these modular structures may be involved in the control (and deregulation) of biological functions found in the disease.

2 Breast Cancer Heterogeneity and Modularity

The study and treatment of breast cancer is complex. Part of this complexity comes from the fact that breast cancer is a heterogeneous disease, which can be seen at the clinical, physiological, and molecular levels. Since the rise of genomic technologies, efforts have been made to find clinical applications of this heterogeneity. One paradigmatic example is the development of molecular classification systems such as PAM50 [10]. Each of these breast cancer *molecular subtypes* (luminal A, luminal B, basal, and HER2-enriched) have differences in prognosis, life expectancy, and pharmacological responses, making them particularly useful in the context of *precision medicine*.

We have shown that transcriptional networks based on MI are able to capture the heterogeneity of breast cancer. In previous work, we have shown that comparable transcriptional networks, derived from gene expression samples classified into different molecular subtypes, exhibit topological differences, including network connectivity and organization. Furthermore, the genes involved in each network (and the corresponding connections between genes) are different, and genes with a high degree in one subtype are not necessarily highly connected in another [6].

Based on the in the connectivity patterns observed in transcriptional networks, and the differences seen between the different molecular subtypes, we decided to explore in [2]: (i) whether transcriptional networks of breast cancer molecular subtypes exhibit modular structures (ii) whether differences in modular structures are observed between subtypes and (iii) whether modules may be associated to known molecular functions.

We found that indeed, there are unique modular structures found in each molecular subtype network. Modules identified in these networks are, in many cases, related to biological functions. The most widely used strategy to identify functional roles is the use of Over Representation Analysis (ORA), in which the overlap between a gene-set of interest, and known gene-sets associated to biological functions (such as those described in the Gene Ontology [4] database) is statistically evaluated, using Fisher's exact (hypergeometric) test, in order to identify what is commonly known as *enrichment*. In this aspect, using modules to identify gene-sets of interest has been shown to be superior to traditional methods of selection, such as lists of over and subexpressed genes, or hierarchically clustered expression genes: module detection relies only on the network

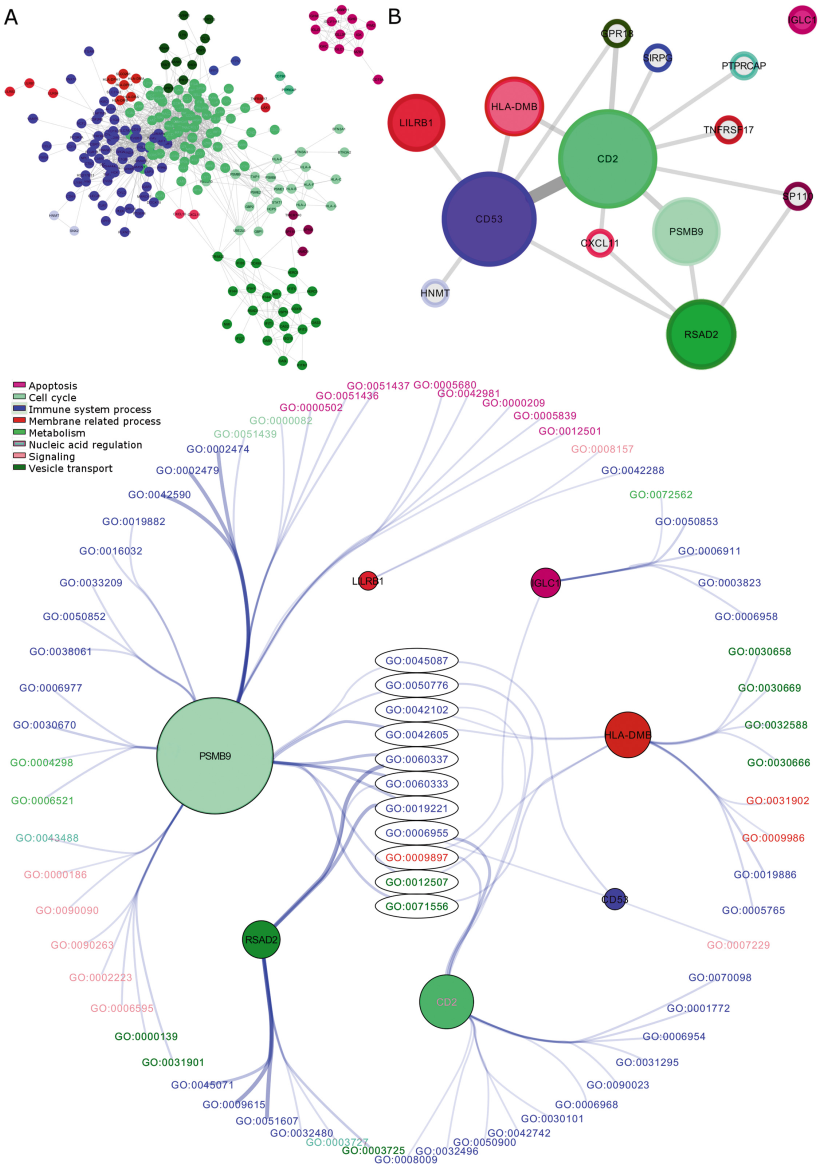


Fig. 1. Modularity in the basal breast cancer network. In panel A, we may observe the transcriptional network for the basal molecular subtype of breast cancer. This network is split into *modules*, each having more links between nodes inside the module than to nodes outside the module; in the figure, each module has a different color. In panel B, a coarse-grained network visualization, where nodes represent modules (labeled with the name of the highest degree gene) and the edges are proportional to the number of links between modules, also representing the information flow between modules. In this visualization, solid nodes are associated to biological functions. Panel C shows the biological functions (indicated using Gene Ontology nomenclature) that are associated to each module; it can be seen that there is small overlap in biological features between modules. *Figure adapted from* [2]

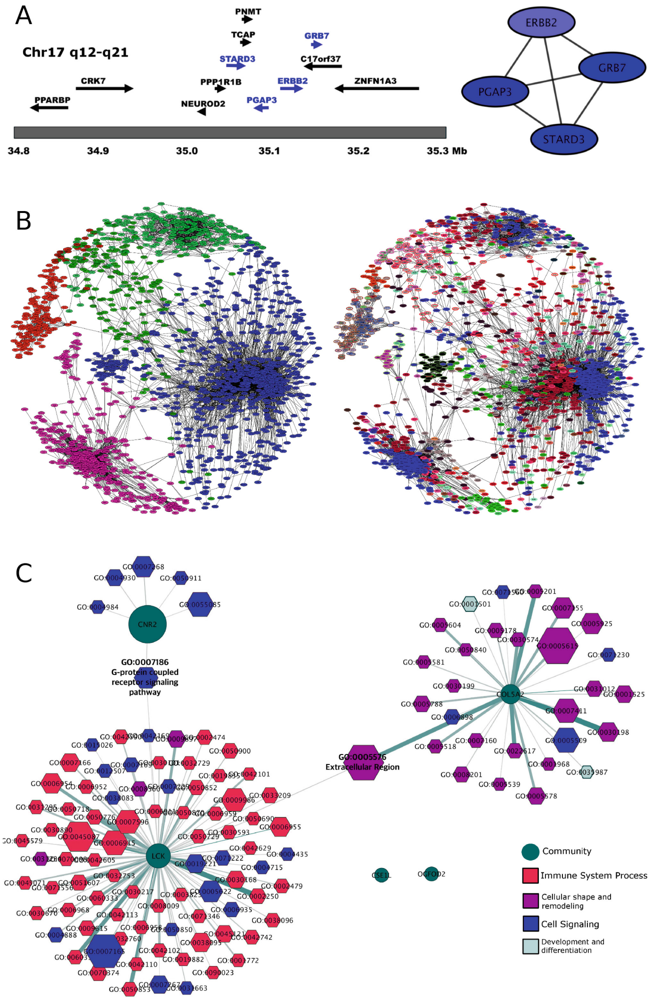


Fig. 2. Hierarchical modularity in the HER2-enriched breast cancer network. In the top panel, the HER2-enriched amplicon, which is a causal agent for the development of HER2-enriched breast cancer, is shown. In subpanel A, the chromosomal ideogram is shown; in subpanel B, the small connected component of the transcriptional network containing these genes is shown. The middle panel shows the giant connected component of the HER2-enriched transcriptional network: subpanel A shows nodes colored by their modules, and subpanel B shows nodes colored by hierarchical submodules, higher definition partitions of the aforementioned modules. Bottom panel shows the association of high-level modules to biological processes. *Figure adapted from* [3]

topology associated to a phenotype (in this case, defined by the co-expression patterns recovered through MI), therefore not needing arbitrary cut-off points or comparisons between phenotypes [1].

We used the ORA approach in order to identify the relationships between the modules in the molecular subtype transcriptional networks and biological functions. Among the results described in the aformentioned publication [2], we would like to highlight the findings related to the basal subtype. In the network for this subtype, we find a large connected component that is comprised of several modules. Most of these modules may be associated through ORA to biological functions; these biological functions are all, in one way or another, to the immune system. These processes had not been described before as exclusively associated to the basal subtype, let alone to a specific set of genes involved in the basal regulatory program. One of these modules, composed by 21 genes including PSMB9 (used as the module identification as it is the gene with the highest degree in the module), is involved not only with immune processes, but also with 30 other processes, including apoptosis. These results are illustrated in Fig. 1. The understanding of the connectivity patterns of apoptosis-related genes may be a starting point for the development of novel therapeutic schemes.

Recognizing that the modular structure of a network may have a hierarchical behavior [13], we explored the hierarchical modularity of the HER2-enriched molecular subtype [3]. We observed how different levels of modular description may capture different biological phenomena occurring at different scales. For instance, we were able to capture the connections between the genes of the HER2-amplicon, which is widely understood to be the genetic aberration responsible for this breast cancer subtype, at the connected component level. First level Infomap modules in subgraphs not connected to the HER2-amplicon were involved in processes such as signaling, immunity and cellular morphology. Higher resolution submodules were associated to more specific functions, such as the regulation of micro-RNA expression and the activation of viral-like immune responses. These results are illustrated in Fig. 2.

These results highlight the use of information theory-based methods for transcriptional network reconstruction and module detection for the study of the heterogeneity of breast cancer. We have shown examples of how using these methods it is possible to identify the modular structures associated to each molecular subtype, and through these modules, associate specific biological functions to these phenotypes.

3 Conclusions

Modularity in networks has proven to be a theoretical foundation for many useful models, allowing the exploration of functional characteristics of biological systems. The use of algorithms to find modules into a biological network may help to understand underlying processes behind the appearance of certain features. The combination of network theory with biological knowledge is mandatory to have a clear idea how the modular architectures shape a specific phenotype.

In this work we stress how different is the modular and submodular structure of breast cancer subtype networks, in terms of the architecture and more importantly, in the functionality of said modules. With these network approaches we have been able to observe several instances of how alterations in modularity are associated to different cancer phenotypes. As an *in silico* model, experiments are needed to corroborate those findings; nonetheless, as new models that use different setups, such as data sets, network inference, and modularity detection algorithms, show consistent results, it seems that Information Theory-based network modularity approaches are robust tools to understand the behavior of the modular structure and functionality in breast cancer, which may in turn lead to important insights for the future of personalized network medicine.

References

1. Alcalá-Corona, S.A., Velázquez-Caldelas, T.E., Espinal-Enríquez, J., Hernández-Lemus, E.: Community structure reveals biologically functional modules in MEF2C transcriptional regulatory network. Front. Physiol. **7**, 184 (2016)
2. Alcalá-Corona, S.A., De Anda Jáuregui, G., Espinal-Enríquez, J., Hernández-Lemus, E.: Network modularity in breast cancer molecular subtypes. Front. Physiol. **8**, 915 (2017)
3. Alcalá-Corona, S.A., De Anda Jáuregui, G., Espinal-Enríquez, J., Hernández-Lemus, E.: The hierarchical network structure of HER2+ breast cancer. Submitted to Front. Physiol. Sect. Syst. Biol. (2018). (in Press)
4. Gene Ontology Consortium. Gene ontology consortium: going forward. Nucleic Acids Res. **43**, D1049–D1056 (2015)
5. de Anda-Jáuregui, G., Mejía-Pedroza, R.A., Espinal-Enríquez, J., Hernández-Lemus, E.: Crosstalk events in the estrogen signaling pathway may affect tamoxifen efficacy in breast cancer molecular subtypes. Comput. Biol. Chem. **59**, 42–54 (2015)
6. de Anda-Jáuregui, G., Velázquez-Caldelas, T.E., Espinal-Enríquez, J., Hernández-Lemus, E.: Transcriptional network architecture of breast cancer molecular subtypes. Front. Physiol. **7**, 568 (2016)
7. Fortunato, S.: Community detection in graphs. Phys. Rep. **486**(3–5), 75–174 (2010)
8. Fortunato, S., Hric, D.: Community detection in networks: a user guide. Phys. Rep. **659**, 1–44 (2016)
9. Margolin, A.A., Wang, K., Lim, W.K., Kustagi, M., Nemenman, I., Califano, A.: Reverse engineering cellular networks. Nat. Protoc. **1**, 662–671 (2006)
10. Parker, J.S., Mullins, M., Cheang, M.C.U., Leung, S., Voduc, D., Vickery, T., Davies, S., Fauron, C., He, X., Hu, Z., et al.: Supervised risk predictor of breast cancer based on intrinsic subtypes. J. Clin. Oncol. **27**(8), 1160 (2009)
11. Ravasz, E., Somera, A.L., Mongru, D.A., Oltvai, Z.N., Barabasi, A.L.: Hierarchical organization of modularity in metabolic networks. Science (New York, N.Y.) **297**, 1551–1555 (2002)
12. Rosvall, M., Bergstrom, C.T.: Maps of random walks on complex networks reveal community structure. Proc. Natl. Acad. Sci. **105**(4), 1118–1123 (2008)
13. Rosvall, M., Bergstrom, C.T.: Mapping change in large networks. PLoS ONE **5**(1), e8694 (2010)
14. Sole, R.V., Valverde, S.: Spontaneous emergence of modularity in cellular networks. J. R. Soc. Interface **5**, 129–133 (2008)
15. Solé, R.V., Valverde, S., Rodriguez-Caso, C.: Modularity in biological networks. In: Biological Networks (2006)
16. Tripathi, S., Moutari, S., Dehmer, M., Emmert-Streib, F.: Comparison of module detection algorithms in protein networks and investigation of the biological meaning of predicted modules. BMC Bioinform. **17**, 129 (2016)

A Manifold Learning Approach to Chart Human Brain Dynamics Using Resting EEG Signals

Hiromichi Suetani[1,2(✉)], Yoko Mizuno[2], and Keiichi Kitajo[2]

[1] Faculty of Science and Technology, Oita University, Oita 870-1192, Japan
suetani@oita-u.ac.jp

[2] Riken Center for Brain Science, RIKEN, Wako, Saitama 351-0198, Japan

Abstract. In this study, we propose an approach to identify individuality that appears in human brain dynamics and visualize inter-individual variations in a low-dimensional space. For this purpose, we first introduce an appropriate similarity measure between multichannel electroencephalography (EEG) signals based on information geometry. Then, we use t-distributed stochastic neighbor embedding, which is a state-of-the-art algorithm for manifold learning, and combine it with the information distance to map points in the high-dimensional EEG signal space into two-dimensional space. We show that a fine low-dimensional visualization enables us to identify each subject as an isolated island of points and preserve inter-individual variations. We also provide an appropriate approach to tune the cost function parameter.

Keywords: EEG · Individuality · Information geometry
Manifold learning · Personal authentication · Spectral analysis

1 Introduction

Human beings exhibit their genetic and habitual traits in their physiological characteristics, including the face, iris, and fingerprints, and behavioral characteristics, such as the voice and gait. The use of such biometrics has recently received attention from the viewpoint of the engineering purpose, for example, automated personal authentication. In addition to appearances, physiological signals in vivo often reflect individuality. For example, researchers have already discussed that there is a relationship between the characteristics of electroencephalography (EEG) signals and genetic factors [1]. Figure 1 shows the time series of multichannel EEG for four subjects from our experiments explained below. Our visual inspection of Fig. 1 indicates that EEG waveforms are different from each other, that is, some degree of "individuality" emerges from such physiological signals.

In this study, we propose an approach based on "manifold learning," which is a general framework of unsupervised nonlinear dimension reduction for visualizing high-dimensional data in a lower-dimensional space [2], to identify each

A. J. Morales et al. (Eds.): ICCS 2018, SPCOM, pp. 359–367, 2018.
https://doi.org/10.1007/978-3-319-96661-8_37

individual and visualize inter-individual variations in human brain dynamics. For this purpose, we analyze the datasets of EEG signals recorded from 100 volunteers while they closed their eyes at rest. We focus on the frequency domain information of EEG signals and provide a reasonable distance measure between EEG power spectral densities (PSDs) based on the theory of information geometry [3]. Using this information distance, we use a modification of t-distributed stochastic neighbor-embedding (t-SNE) [4], which is a state-of-the-art algorithm for manifold learning. We show that our proposed method is useful for identifying each individual from other subjects sufficiently. We also discuss how a low-dimensional "chart" concerned with human brain dynamics can be used to capture inter-individual variations.

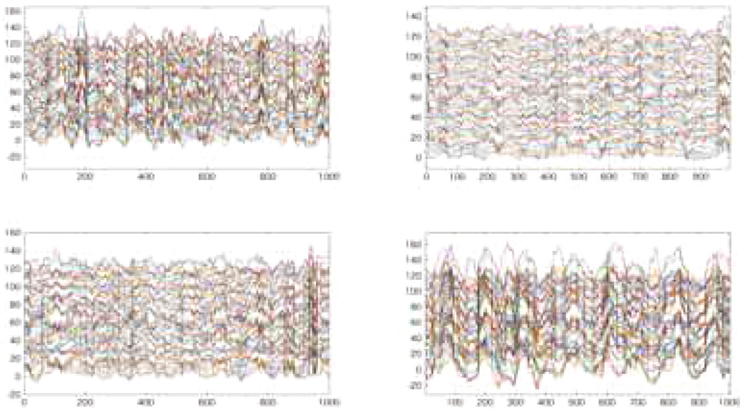


Fig. 1. Sixty-three ch EEG signals for four subjects.

2 Preparation of Datasets

2.1 Measurements of EEG Signals

In this study, a total of 100 healthy volunteers (male: 54, female: 46) participated in our experiments after giving informed consent that was approved by the ethics committee of RIKEN. We measured multiple EEG signals at a sampling rate of 1000 Hz while participants closed their eyes for 3 min. An EEG amplifier (BrainAmp MR+, Brain Products GmbH, Germany) with a 63-ch EEG cap (Easycap, EASYCAP GmbH, Germany) was used in the measurements. Electrodes were positioned according to the international 10/10 system with a left earlobe reference and Afz as the ground. EEG signals were offline re-referenced to the average of the left and right earlobe references and bandpass filtered (FIR filter) between 3 Hz to 80 Hz. Then, we applied an independent component analysis (ICA)-based automatic artifact removal procedure to remove generic artifacts, and eye movements and blink-related artifacts.

2.2 Calculations of Power Spectral Densities (PSDs)

After preprocessing the EEG data, we performed Fourier analysis on the datasets to obtain the frequency domain information of EEG. First, for each subject, the time series with the duration of 180 s was divided into 18 non-overlapping segments (i.e., the time length of each segment was 10 s). Therefore, there were 18 segments of the 63-dimensional EEG time series for each subject. Then, for the EEG time series with 10 s of each channel, we performed fast Fourier transformation (FFT) and obtained a power spectrum density (PSD) by taking the average of PSDs over 10 1-s sub-segments that did not overlap. Because there were 100 subjects, a total of 1800 datasets with 63 PSD dimensions was the target of the analysis in this study.

3 Analysis Methodology

3.1 Distance Measure Based on Itakura–Saito Divergence

In the field of data science, it is essential how to quantify similarities or dissimilarities (i.e., "distances") between items of data. For a dataset that consists of density functions, the information-theoretic divergence, such as the Kullback–Leibler, Hellinger, and Jensen-Shannon divergences, is suitable for measuring similarities between probability distributions rather than the Euclidean squared distance. Particularly, from the theory of information geometry that induces the Riemannian geometrical structure to a set of probability distributions [3], the Kullback–Leibler divergence between two PSDs can be calculated as

$$D_{\mathrm{IS}}(S_1\|S_2) = \int (\frac{S_1(f)}{S_2(f)} - \log\frac{S_1(f)}{S_2(f)} - 1)\mathrm{d}f. \tag{1}$$

Eq. (1) is in agreement with the "Itakura–Saito divergence" that was originally proposed by Itakura and Saito from the maximum likelihood estimation of short-time speech spectra under autoregressive modeling, and is popular in the speech community [5].

Because Itakura–Saito divergence is not a true metric, that is, not symmetric, we consider the arithmetic mean value as $\{D(S_1\|S_2) + D(S_2\|S_1)\}/2$. Additionally, there are 63 PSDs for each EEG segment; therefore, we introduce the following distance measure between the n-th and m-th items as

$$d_{nm} = \frac{1}{63}\sum_{l=1}^{63}\frac{1}{2}\{D_{\mathrm{IS}}(S_n^{(l)}\|S_m^{(l)}) + D_{\mathrm{IS}}(S_m^{(l)}\|S_n^{(l)})\}, \tag{2}$$

where $S_{n(m)}^{(l)}(f)$ denotes the PSD of the l-th electrode channel of the $n(m)$-th dataset.

3.2 Manifold Learning

Classical Scaling: Multidimensional scaling (MDS) is a generic term of the dimensionality reduction of data when information on the pairwise dissimilarities (distances) between items is available [2]. Particularly, when the dissimilarity is measured using the Euclidean distance, the corresponding linear version of MDS is called classical scaling; that is, classical scaling determines the linear mapping that minimizes cost function $J = \sum_{nm}(e_{nm}^2 - \|\boldsymbol{y}_n - \boldsymbol{y}_m\|^2)$, where e_{nm} denotes the Euclidean distance between the n-th and m-th items in the original high-dimensional space, and $\|\boldsymbol{y}_n - \boldsymbol{y}_m\|$ denotes that between the embedding points of these two items in the low-dimensional space. The minimization of cost function J is solved by the eigen-decomposition of Gram matrix $K = XX^t$, where element k_{nm} of matrix K is the double-centering of squared Euclidean distance e_{nm}^2. Therefore, classical scaling is equivalent to performing principal component analysis on Gram matrix K.

Isometric Feature Mapping (ISOMAP): ISOMAP, which was proposed by Tenebaum et al. [6], is a generalization of classical scaling to nonlinear manifolds. It is based on replacing the Euclidean distance by an approximation of the geodesic distance on the manifold. In this approximation, for each data point on manifold $\mathcal{M}$, k (or a radius of ϵ) neighborhood points in the sense of the Euclidean distance are connected by edges as a local graph. By repeating this process for all points, global graph $\mathcal{G}$ that covers $\mathcal{M}$ is constructed. Then, the geodesic distance between two points is approximated as the length of the shortest path that connects these two points on $\mathcal{G}$ using Dijkstra's algorithm. Finally, by conducting the process in the same manner as classical scaling using the approximated geodesic distance matrix, ISOMAP is accomplished.

***t*-Distributed Stochastic Neighbor Embedding (*t*-SNE) and Its Modification:** Hinton and Roweis proposed a version of MDS called stochastic neighbor embedding (SNE) [7] that preserves neighbor relations between items in the original high-dimensional space in the lower-dimensional space using the notion of "neighbor probabilities," which plays the role of similarities between items.

Given a set of N items $\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_N$ in the original high-dimensional space, SNE first calculates

$$p_{m|n} = \frac{\exp(-\|\boldsymbol{x}_n - \boldsymbol{x}_m\|^2/2\sigma_n^2)}{\sum_{r \neq n} \exp(-\|\boldsymbol{x}_r - \boldsymbol{x}_n\|^2/2\sigma_n^2)}, \tag{3}$$

where $\|\boldsymbol{x}_n - \boldsymbol{x}_m\|$ typically uses the Euclidean distance. Additionally, bandwidth σ_n of the Gaussian kernel is calculated using a binary search method that ensures that the entropy of the conditional probability $p_{m|n}$ as a function of m is approximately equal to $\log(k)$, where k is the effective number of neighbors and called the "perplexity" parameter. Conditional probability $p_{m|n}$ means that $\boldsymbol{x}_m$ is picked up as a neighbor of $\boldsymbol{x}_n$ proportional to the Gaussian probability centered at $\boldsymbol{x}_n$. The values of $p_{m|n}$ and $p_{n|m}$ are not necessarily equivalent because σ_n and σ_m are set to different values. Thus, by also considering $p_{n|m}$ in a similar manner, the neighbor probability between items $\boldsymbol{x}_n$ and $\boldsymbol{x}_m$ can be obtained as $p_{nm} = (p_{m|n} + p_{n|m})/2$.

SNE seeks a set of N items $\boldsymbol{y}_1, \boldsymbol{y}_2, \cdots, \boldsymbol{y}_N$ in a lower-dimensional embedding space that preserves similarities p_{nm} in the original space as much as possible. To achieve this, we also consider the neighbor probability between two items $\boldsymbol{y}_n$ and $\boldsymbol{y}_m$ in the embedding space as

$$q_{nm} = \frac{\exp(-\|\boldsymbol{y}_n - \boldsymbol{y}_m\|^2)}{\sum_{r \neq n} \exp(-\|\boldsymbol{y}_r - \boldsymbol{y}_n\|^2)}. \tag{4}$$

Finally, the configuration of $\boldsymbol{y}_n$ is obtained by minimizing the sum of the Kullback–Leibler divergences between two distributions $P_n = \{p_{nm} | m = 1, 2, \cdots, N\}$ and $Q_n = \{q_{nm} | m = 1, 2, \cdots, N\}$ as

$$\mathcal{C} = \sum_n KL(P_n \| Q_n) = \sum_n \sum_{m \neq n} p_{nm} \log \frac{p_{nm}}{q_{nm}}. \tag{5}$$

The minimization of the cost function Eq. (5) with respect to the configuration of $\boldsymbol{y}_n$ is performed using the gradient descent method.

In practical applications, it is difficult to minimize Eq. (5) when the Gaussian distribution is taken in Eq. (4). In [4], Maaten and Hinton improved the original SNE by considering a heavy-tailed *Student's-t* distribution:

$$\tilde{q}_{nm} = \frac{(1 + \|\boldsymbol{y}_n - \boldsymbol{y}_m\|^2)^{-1}}{\sum_{r \neq n} (1 + \|\boldsymbol{y}_r - \boldsymbol{y}_n\|^2)^{-1}}, \tag{6}$$

which is used to measure the similarities between items in the lower-dimensional space. This improvement of SNE is called t-SNE.

Although the conventional (t-)SNE algorithm considers the Euclidean distance as the metric, its choice is arbitrary. Particularly, because the Euclidean distance is affected by the curse of dimensionality in Eq. (3), we consider an alternative in an appropriate manner. Therefore, we consider the distance measure defined in Eq. (2) in this study.

4 Results

Figure 2 shows the results of our experiments with classical scaling, ISOMAP, and t-SNE based on Itakura–Saito divergence between the PSDs of EEG for 100 subjects. Each filled circle point denotes a single item that corresponds to a set of 63 PSDs concerned with an EEG segment for a single subject; there are 18 points for a subject. Additionally, each color and number represents each subject.

Solutions produced using classical scaling (Fig. 2(a)) and ISOMAP (Fig. 2(b)) demonstrate large overlaps between the subject classes. In particular, in Fig. 2(a), subjects' data points are densely mixed around the center of the chart, which means that classical scaling as a linear dimensionality reduction method is

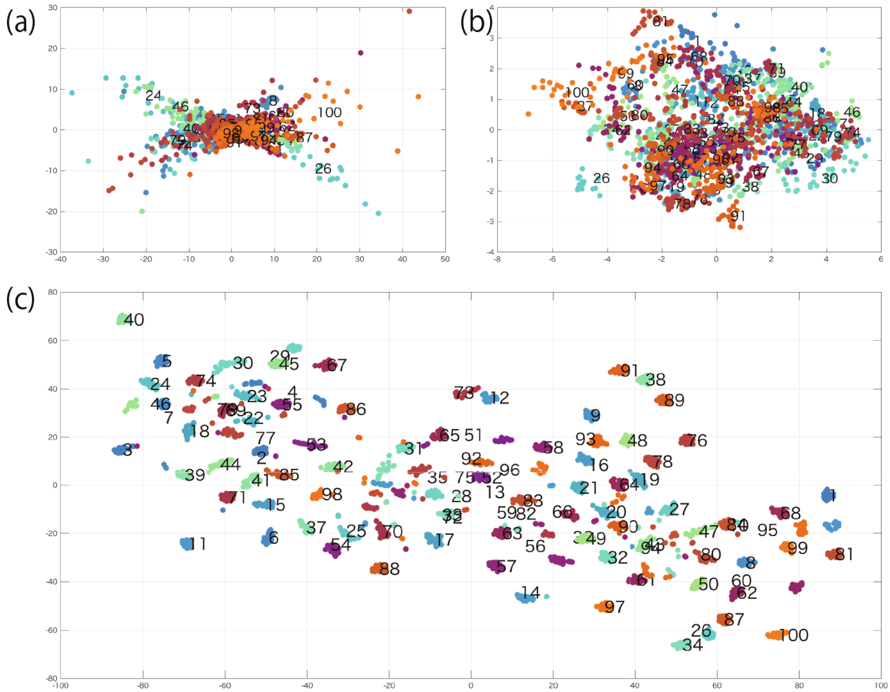


Fig. 2. Two-dimensional visualizations of EEG segments with (a) classical scaling, (b) ISOMAP ($k = 20$), and (c) t-SNE ($k = 30$).

not suitable for visualizing the high-dimensional feature of the PSDs in two-dimensional (2D) space. In Fig. 2(b), which is the case of ISOMAP, there are improvements compared with the case of classical scaling, that is, there are some separations between subjects' classes, but they are insufficient. By contrast, Fig. 2(c) shows that t-SNE constructs a chart in which the separations between subjects' classes are almost perfect, that is, a set of points for a single subject creates well-separated local "islands."

Furthermore, t-SNE preserves not only local (individual) information but also global (inter-individuals) information about the PSDs in the low-dimensional space. Figure 3 shows the comparison between the distance matrix based on Itakura–Saito divergence in the original high-dimensional PSDs space and that based on the Euclidean distance in the 2D space constructed using t-SNE. We see that similarities not only between the same subject's points but also between different subjects' points (the off-diagonal part of the matrix) are held in the t-SNE chart.

The result of t-SNE, however, strongly depends on the choice of cost function parameters. In the case of t-SNE, the perplexity that controls the effective number of neighborhoods affects the visualization. Figure 4(a) shows the visualizations using t-SNE for four values of perplexity k. Generally, when we choose a small value of k, the corresponding solution emphasizes the local structure.

In the case of $k = 10$, although islands are more isolated compared with the case of $k = 30$, the global structure becomes isotropic (circle-like), which means that the similarities between different subjects in the original space do not hold in this low-dimensional space. By contrast, when we use a large value of k, for example $k = 290$, islands of points expand, and they are mixed; that is, depending on the choice of their perplexity, there is a trade-off between the ability of the individuality identification and that of the inter-individual variation.

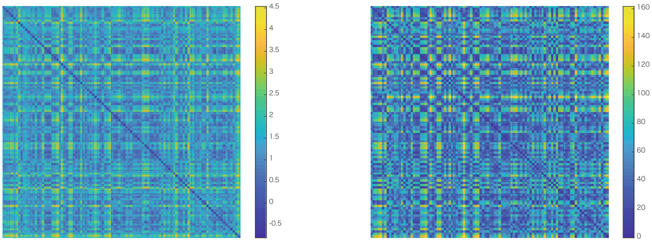


Fig. 3. The Itakura–Saito divergence-based distance matrix in the original PSDs space (left, the logarithmic scale) and the Euclidean distance matrix in the t-SNE space (right).

Therefore, we introduce two indices that evaluate the performance of t-SNE. First, index G is concerned with the global structure, that is, how the Euclidean distance matrix between items in the t-SNE space is similar to that in the original PSDs space, which is based on Itakura–Saito divergence. Such an index G is defined as

$$G = \frac{|(1/N_*)\sum_{n=1}^{N-1}\sum_{n<m}(d_{nm}-\bar{d})(e_{nm}-\bar{e})|}{\sqrt{(1/N_*)\sum_{n=1}^{N-1}\sum_{n<m}(d_{nm}-\bar{d})^2}\sqrt{(1/N_*)\sum_{n=1}^{N-1}\sum_{n<m}(e_{nm}-\bar{e})^2}}, \quad (7)$$

where $N_* = N(N-1)/2$, d_{nm} is the distance measure between two PSDs defined in Eq. (2), e_{nm} is the Euclidean distance between items in the t-SNE space, and $\bar{d}$ and $\bar{e}$ are their arithmetic mean values. Second, index L is concerned with the local structure, that is, how an island of points that belongs to a single subject is well separated from other islands of points in the t-SNE space. According to the framework of Fisher's linear discriminant analysis [2], such an index can be defined as the ratio of the variance between the classes to the average of the variance within the classes.

$$L = \frac{\det(\mathrm{cov}(\boldsymbol{c}^{(n)}))}{(1/N)\sum_{n=1}^{N}\det(\mathrm{cov}(\boldsymbol{u}_j^{(n)}))}, \quad (8)$$

where $\boldsymbol{u}_j^{(n)}, j = 1, 2, \cdots, 18$ denotes points in the t-SNE space that belong to the island of the n-th subject and $\boldsymbol{c}^{(n)}$ denotes the centers of them, respectively. The lower panel of Fig. 4(b) shows the product of G and L as a function of k.

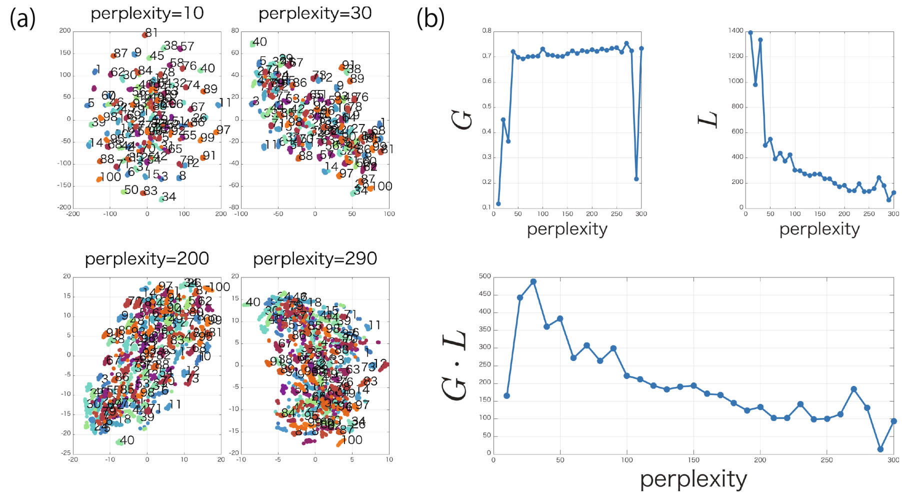


Fig. 4. (a) t-SNE 2D visualizations of EEG segments for four different values of the perplexity parameter. (b) The global (G) and local (L) structure indices as functions of the perplexity (upper panels). The product $G \cdot L$ as a function of the perplexity (lower panels).

We see that there is an optimal value of the perplexity at around $k = 30$ that provides an optimal visualization in the sense that the local identifiability of individuals and the global structure of inter-individual variations are balanced in the low-dimensional space.

5 Conclusion

We conclude that the proposed approach based on manifold learning, which is a framework of unsupervised learning methods for nonlinear dimensionality reduction, successfully identifies individuality in EEG signals. We introduced an appropriate distance based on Itakura–Saito divergence that has the rigid theoretical background of information geometry to measure the similarity between the PSDs of EEG signals. The combination of this information distance measure with t-SNE, which is a state-of-the-art manifold learning algorithm, enables us not only to visualize different subjects as isolated islands of points in 2D space but also to arrange such islands with appropriate locations according to the similarities in the original PSDs without any supervised information. The proposed approach in this study can also incorporate other characteristics of human brain dynamics, such as functional connectivity measures (e.g., cross spectra and phase synchronization) by applying other types of information distance. Such further investigations will provide deeper interpretations of charts concerned with human brain dynamics in healthy and diseased conditions, and will possibly lead to the diagnosis of neurological and psychiatric disorders.

Acknowledgements. This study was supported by "Actualize Energetic Life by Creating Brain Information Industries," ImPACT Program of Council for Science, Technology and Innovation (Cabinet Office, Government of Japan).

References

1. Vogel, F.: Genetics and the Electroencephalogram. Springer, Heidelberg (2000)
2. Hastie, T., Tibshirani, R., Friedman, J.: The Elements of Statistical Learning: Data Mining, Inference, and Prediction, 2nd edn. Springer, Heidelberg (2008)
3. Amari, S., Nagaoka, H.: Methods of information geometry. Am. Math. Soc. **191**, 206 (2007)
4. Maaten, L., Hinton, G.: Visualizing data using t-SNE. J. Mach. Learn. Res. **9**, 2579–2605 (2008). We use codes in Matlab Toolbox for Dimensionality Reduction (https://lvdmaaten.github.io/drtoolbox/) by Ph.D. L van der Maaten
5. Itakura, F., Saito, S.: Analysis synthesis telephony based on the maximum likelihood method. In: Proceedings of the 6th International Congress on Acoustics, pp. C17 –C20 (1968)
6. Tenenbaum, J.B., De Silva, V., Langford, J.C.: A global geometric framework for nonlinear dimensionality reduction. Science **290**(5500), 2319–2323 (2000)
7. Hinton, G.E., Roweis, S.T.: Stochastic neighbor embedding. In: Advances in NIPS, pp. 857–864 (2003)

The Design for Maritime Singularity: Exploration of Human/AI Teaming and Organizational Carrying Capacity for the U.S. Navy

Matthew Largent[1](✉), Garth Jensen[2], and Rebecca Law[3]

[1] SPAWAR Systems Center Atlantic,
P.O. Box 190022, N Charleston, SC 29419-9022, USA
Matthew.Largent@navy.mil
[2] NSWC Carderock Division, 9500 MacArthur Blvd,
West Bethesda, MD 20817-5700, USA
[3] Naval Postgraduate School, 1 University Circle, Monterey, CA 93943, USA

Abstract. The "Design for Maritime Singularity" was a crowdsourcing effort focused on two looming concerns that the U.S. Navy will have to address in the coming decades. The first concern deals with the concept of the technological singularity, where the premise is that Artificial Intelligence (AI) could rapidly expand in capability past human intelligence. This has been hypothesized by some to mean that the AI would leave humanity behind, possibly making humankind irrelevant. The second concern deals with organizational complexity, building off the argument made by Dr. Yaneer Bar-Yam that hierarchical organizations are limited in their ability to handle complexity by the carrying capacity of the small number of individuals who make decisions at the top. As the environment becomes more complex, hierarchical organizations like the U.S. Navy will find themselves in a state where their traditional construct could hinder their ability to process this complexity, thus limiting their effectiveness.

The authors, sponsored by the U.S. Office of Naval Research, created a study to explore these concerns using the collective intelligence platform mmowgli (Massively Multiplayer Online War Game Leveraging the Internet). Over a one-week timeframe players from all over the world collaborated, developing concepts that explored how the U.S. Navy could address the concerns. The themes that emerged from that event paint a picture of how the Navy might adjust so that it could ride the wave of technological change and increasing complexity instead of being swamped. This paper describes the mmowgli game, the themes that emerged from the game, and three more detailed concepts fleshed out in a follow-on workshop.

Keywords: mmowgli · Collective intelligence · Artificial intelligence
AI · Wargame · Singularity · Human machine teaming
Hierarchical organization · Organizational complexity

Sponsor: Work performed supporting the Office of Naval Research, Director of Disruptive Technologies, Dr. Eric Gulovsen, Under Funding Documents: N0001414WX01557, N0001414WX01558, and N0001416WX01726.

A. J. Morales et al. (Eds.): ICCS 2018, SPCOM, pp. 368–379, 2018.
https://doi.org/10.1007/978-3-319-96661-8_38

1 mmowgli Overview

The Massively Multiplayer Online War Game Leveraging the Internet (mmowgli) is a re-purposeable online platform as well as a transformational practice that harnesses the creative potential of large, diverse groups for thinking and acting on complex, open-ended challenges and opportunities. mmowgli also builds the creative capacity of the crowd by offering a more gameful, more novel, and more democratized way of engaging together with a problem than traditional methods afford.

In a typical mmowgli game players are first presented with a Call to Action. The intention is to place the players in a scenario that is narratively incomplete, ending with a broad question posed to the players as the purpose for the game. This immersion in scenario is borrowed conceptually from the futuring work of Schwartz [1] and others and allows the players a creative whitespace in which to form their concepts.

Once the players have immersed themselves in the Call to Action they enter the first phase of the game called "Cardplay". The broad question of the Call to Action is divided up into two more specific components called the Yin and Yang. As referenced in a paper written for the White House Office of Science and Technology Policy (OSTP) [2] the concept of mmowgli cardplay is built upon David Bohm's concepts of dialogue as described in *The Fifth Discipline* [3] by Peter Senge. Ideas represented in mmowgli are built in a dialogue format with initial responses to the Yin and Yang played as 140 character or less seed ideas. Subsequent players then comment on those initial ideas (also in 140 characters or less) in ever growing trees of conversation. These comments are in one of four categories, Expand, Counter, Adapt, and Explore. This is the divergent phase of the game and the goal is to create many complex conversations without limiting or grading ideas.

The last phase of the game is the convergent "Action Plan" phase. Of the hundreds or thousands of conversational threads emerging from Cardplay a relatively small number, typically on the order of 1%, have sufficient critical mass of player energy and/or sufficiently promising ideas that they cross the threshold to become Action Plans. The cards chosen to become action plans are selected by game administers (gamemasters) or self-selected by players. For each action plan, small groups of self-selecting players, many of whom were involved in the original generation and development of the thread during Cardplay, come together as a team online to craft detailed responses to the basic "Who-What-When-Where-Why" questions that help turn a promising idea into a first order plan. For most mmowgli games, the Action Plan phase constitutes the raw output that is turned into concrete recommendations to sponsors and policy makers after the game.

2 Design for Maritime Singularity Game Concept

The "Design for Maritime Singularity" game concept plays intentionally off the wording from the U.S. Navy's Chief of Naval Operations (CNO) "Design for Maritime Superiority" strategy document [4]. At the highest level of abstraction, the game positions players in a future where the long awaited Singularity has arrived, and solicits their collective response to the broad question: "How Might We Design (or re-design) the Navy in light of this Singularity?" The Call to Action was created in the form of a fictional Ted-talk like event [5]. At the more concrete level, the game's Call-To-Action decomposes this broad question into two parts (the Yin and Yang referred to above).

Part one characterizes the "Singularity" in the way that people commonly understand it, as the emergence of greater than human intelligence from technological means, a proposition attributed to scientist and futurist Kurzweil [6]. The emergence of intelligent machines, capable of designing even smarter machines, would create its own kind of Event Horizon (hence the term Singularity), a world in which the unaided human is no longer sufficient and may not even be relevant. Rather than give in this dystopian view the mmowgli narrative cites the invention of freestyle chess, influenced by Kasparov [7] following his defeat by IBM's Big Blue, as a metaphor for how the Navy might approach the pending Singularity. Part One finishes with the following open ended prompt to players (the Yin): "What concepts for human-machine teaming might we develop as we approach Singularity 1?".

Part two introduces an alternative way of thinking about the Singularity (called Singularity 2), drawn from the work of Yaneer Bar-Yam, at the New England Complex Systems Institute [8]. This view argues that a similar phenomenon is happening with environmental complexity as with machine intelligence, namely that complexity is rising to the point that it outstrips individual human ability to manage it. It makes the further argument that organizations, to the extent that they are hierarchical in nature, are limited in their "complexity carrying capacity" by the carrying capacity of the relatively small number of individuals who make decisions at the top of the hierarchy. Thus, traditional organizational constructs are limited in their ability to process this complexity. Part two finishes with the following open ended prompt (the Yang): "As complexity rises all around us, what new organizational constructs should we consider?".

3 Game Execution

The Design for Maritime Singularity mmowgli game ran from March 27th through April 2nd 2017. In that time 1272 players registered and 390 players participated through playing a total of 9109 cards and creating 45 action plans. As is typical for these games the greatest participation was on the first day, with half of the cards played that day. Looking at the Yin and Yang, 609 unique ideas (seed cards) were played in response to the Part one question and 389 seed cards were played in response to the Part two question. Players were allowed to play cards through March 31st and then allowed to continue to work on action plans only for April 1st and 2nd.

The action plan phase of the game started almost immediately, with the first action plan created within two hours of starting the game. It is currently not possible to track the total number of edits on an individual action plan, but there were 1852 comments and 547 author-to-author messages. This level of communication is impressive and represents a high level of collaboration.

4 Analysis of Results

With Design for Maritime Singularity the mmowgli team, with support from the sponsor ONR, piloted a new approach to post game analysis. Traditionally the mmowgli output, consisting of the combined body of Card Play and Action Plans, is made available to the game sponsor who performs, usually with some assistance from team mmowgli, the bulk of analysis, reporting and recommendation writing. With Design for Maritime Singularity the authors inserted an intermediate step, convening a three-day workshop using methods from Human Centered Design. The next section describes the workshop itself. This section describes the process of curating and converting the raw output from the mmowgli game into a form that could be utilized by the workshop participants.

Between the end of the game (April 2, 2017) and the beginning of the workshop (August 15, 2017) the authors combed through all 9109 cards and 45 action plans using an adapted form of hermeneutics, a process more commonly associated with theology and the liberal arts than with wargaming or crowdsourcing. In effect, the authors treated the mmowgli output as a form of literary text, reading it several times through, highlighting major themes as well as significant outliers, occasionally stitching together disjoint threads to form a completely new idea.

The themes are listed below. Where mmowgli raw data is cited, AP followed by a number indicates Action Plan #X. For example AP 34 = Action Plan #34. Otherwise, if a simple number is listed it indicates the Card number from the game, so “7162” means Card #7162.

4.1 Singularity 1 Major Themes

AI for Intelligence Analysis: This set of concepts encompasses an area where there is a great deal of current research, though there is certainly more that can be done. As data from intelligence systems as well as public sources increases and becomes more complex humans will need help turning that data into information and knowledge. Much of the current work in this area involves machine learning.

Swarming: The concept of swarms often is considered in terms of robots working independently to accomplish a task. In the case of some of these concepts, however, the players look at swarms of manned and unmanned assets working together and also at the idea that a swarm of humans could support or be supported by a single AI. As sensors and computers become smaller and more capable the idea of having inexpensive swarms becomes less of an idea and more an expected reality.

Health-Related Concepts: This is just a small number of concepts, and focuses on how AI or robots could help humans to heal or get treatment faster in battlefield scenarios as well as concepts dealing with using AI to help with medical advances.

Tactical Application of AI to Warfighting: This concept has two main areas. In the first the AI is fighting war through offensive operations or defensive operations. In the second area are concepts that explicitly look at how to attack AI. These range from EMP bursts to concepts based around deception or destroying human trust in the AI.

Creation of AI: This theme deals with different ways to create and train AI. Some are conceptual, like the idea of Moravec's paradox, but most are more practical dealing with creation of hardware and software that will assist with advanced AI.

Transhumanism: This group of concepts takes the basis of the game, which was the merging of AI and human capability, and assumes a physical merger. It involves creating cyborgs, humans with machine elements that allow the merged entity to operate more effectively than either part could operate separately.

4.2 Singularity 1 Design Challenges

Design Challenge: AI Personal Assistant: This set of concepts was focused on personal Artificial Intelligence aids that were created with either an individual or a specific job/rate in mind. These assistants would be in the vein of Siri or Alexa, but much more competent. For some of the concepts the assistant would be very specific to a certain task or set of tasks while other concepts envisioned an assistant that would help with many different topics. One of the most interesting elements of this series of ideas is the personalization aspect, that through dedicated coders or through learning algorithms the AI was customized for an individual, not a one-size-fits-all solution. Represented by mmowgli raw data: *Personal assistant learns how you work and is better able to help – AP3, Developers code solutions alongside sailors that use them, AI developed for specific rates – AP9, AI personal assistant that recognizes and shares best practices – AP39, Technology/AI as a colleague to navigating workplace complexity – AP40, AI grows with you over service time, stays with you as your career progresses – 11, 816, Partner AI with person/thing it will mimic, allow it to grow with partner – 1363, Increased trust in AI if it is paired with you – 279, AI Amanuensis (butler) – 1906, AI advisor for grunt/sailor – 1965, 2645, 5192, AI teaching human – 3950, 6748, AI can increase the capacity for humans to handle complexity – 2142.*

Design Challenge: Interface Between Humans and Computers/Machines/AI: This set of concepts is all about how the human and the computer can communicate with each other. Some look at ways to teach the computer to understand humans, these deal in part with maturing areas of research like text and speech analytics. Another sub-set looks at the idea that humans need to speak/communicate with more precision and therefore the changes should be the humans becoming more computer-like. The third area focuses on more direct connections between the human computer (our brain) and the machine, through direct brain interfaces, EEG (Electroencephalogram) interfaces, and other means. A last element is providing something like Augmented Reality (AR) or Virtual Reality (VR) or even a suit that overlays the computer information into the human's world and connects the human's data to the computer's world. What isn't necessarily discussed is the cognitive load of the additional information these connections will give. Represented by mmowgli raw data: *Natural language and other comfortable interactions with computers – AP11, train people how computers think – AP14, Intelligent suit* *that connects you to network and provides compute, AI, sensing – AP15, Armor to help with medical issues – AP16, incorporate computer programming elements into human language – AP36, Direct brain-machine interface – 13, 15, 469, 568 (brainwaves), 2053(brainwaves), 6056 (telepathy), 6315 (EEG), 6391, Human/AI Application Program Interface (API) – 51, Common language for human/AI communication – 538, 4891, 6267, VR/AR interface – 903, 4332, Attack other people's human/AI interface – 7239.*

Design Challenge: AI Decision Aids: This is a broad range of concepts that approach the idea of how decisions will be made as humans and machines work more closely together. Some of these concepts have to do with the structures within which we make decisions, such as decentralized decision making. Some concepts are about having the AI help with small decisions. Some look at how AI might break up our traditional decision making process, with the AI being the Commanding Officer (CO), the AI breaking the chain of command, or the AI being a red team to point out human failings. Represented by mmowgli raw data: *stock market for ideas related to strategic concepts – AP13, mmowgli helps promote transparency – AP18, push decision making to swarms (mostly S2) – AP28, Use Agile as structure for decision making (mostly S2) – AP29, How did the machine make the decision? – 50, Learn from past decisions – 86, Allowing the machine to decide for the human – 137, Decisions at machine speed – 328, If organization is decentralized, how does AI factor into C2/making decision – 611, AI point out human bias or play devil's advocate – 619, 774, Machine making the human decision easier, making lower level decisions to free up human for higher level – 987, AI break the chain of command – 1112, 2728, 4807, AI as CO – 2161, Multiple decision aides (like have multiple staff members) – 4357, 4955, AI as too easy to predict? – 4991, Train AI to run Prediction markets – 6556.*

These themes represent concepts that the U.S. Navy can use in directing research towards human/machine or human/AI teaming. Additionally, the U.S. Navy can benefit from the wide range of existing research and development already ongoing in these areas.

4.3 Singularity 2 Major Themes

The Role of Surprise, Imagination and Novelty vs the Role of Analysis and Logic: This theme addresses the question of whether an entity (i.e. either human or AI) could ever "analyze" its way into something as imaginative and effective as Doolittle's Raid.

Naïve Rationalism: This theme revolves around the assumption that machine driven, "perfect" algorithmic logic, applied to human affairs, would result in superior outcomes.

Projecting Human Social Complexity onto AI: Many threads questioned whether AI would manifest as a single, monolithic, all-knowing entity, or whether it would be balkanized. In an ironic twist for a game devoted to taming complexity, this led to the need for a proliferation of distinct AI's whose purpose was to keep other AI's in check, creating a situation where the artificial world, rather than helping humanity tame complexity, actually contributes to increasing complexity.

The Role of Feelings/Emotions: Emotions and feelings are one way that humans curate overwhelming amounts of raw, complex data coming in from the world. Additionally, emotions form a key, underappreciated part of the human decision making process. This theme revolves around the role of feelings and emotions in AI.

The Role of Embodied and Tacit Knowledge: Similar to feelings and emotions, humans also possess tacit and embodied knowledge. Is there an analog to tacit/embodied knowledge for AI? Can AI possess embodied/tacit knowledge? If so, how would that manifest? Are we assuming that the vast majority of knowledge is explicit knowledge? If so, what are the implications for AI?

4.4 Singularity 2 Design Challenge

Design Challenge: Treating the U.S. Navy as a Complex Adaptive System: Going back to the Call-To-Action, the goal was to redesign our Navy's organizational construct, at any level, such that we would have a Navy that is robust in any environment, and able to deliver effects matching the scale and complexity of the situation at hand. For the purposes of the design challenge given to the workshop, "at any level" could mean at the large organizational level, e.g. at the Requirements Setting and Resource Sponsor level; it could apply to the broad enterprise that manages research, development and innovation; or it could mean at the operational level, whether an individual unit, or squadron. The primary challenge for this group was to settle on a specific or narrow enough idea to further develop, and to also settle on a specific level of the large Navy organization on which to apply the idea. Represented by MMOWGLI raw data: *Treating the Navy as a Complex Adaptive System. – AP7, How Might the Navy's organizational construct, at any or all levels, need to change in order to push decision making and problem solving to swarms? – AP28, Test and apply Agile Methodology/SCRUM/KANBAN to Navy organizational constructs. – AP29, If the Navy evolves to a more complex, less hierarchical structure, what incentives might emerge to replace traditional hierarchical/bureaucratic incentives? – AP30, Pre-empting the Third Singularity: (note: the third Singularity, refers to that point in time when the Defense top line budget intersects with the increasing per-unit-cost of a given platform, resulting in a Navy force structure consisting of exactly one very large, but very capable, platform). – AP37, Organic structure where the resources would be redirected to address issues as they arise, like the body fights illness or injuries. – 83, For major acquisitions use AI and big data to automate requirements identification and validation. – 99, Why are we assuming the individual carrying capacity for complexity is fixed? – 2142, Shift the burden of policy enforcement from humans to AI. – 2404, Complex, Adaptive Enterprise. Extend the study of complexity and complex adaptive systems to large organizations and enterprises like DoD. – 3516, Can we measure, in real time, if an organizational structure deals with complexity well? – 4448, Human subjectivity in law/policy enforcement is a nightmare. Can we create a system that relies entirely on AI and past results? – 6031.*

5 Workshop Design and Results

In parallel with the qualitative data analysis described in Sect. 4 the authors worked with the Consortium for Robotics and Unmanned Systems Education and Research (CRUSER) Program at the Naval Postgraduate School (NPS) to plan and conduct a workshop. The purpose of the workshop was to focus on the identified design challenges and flesh them out more fully, to the point that they would contain actionable recommendations. The format chosen for the workshop was to conduct a design sprint, over three days, using design methods and principles.

The authors invited a pool of 24 participants drawn from the Naval Research and Development Enterprise (NR&DE), from other government agencies, and from the player pool. After a plenary session where participants were introduced to the game themes and design challenges participants were split into three facilitated working groups. NPS supplied the individual facilitators for each group, as well as an overarching workshop/roving facilitator. Two groups were assigned to Singularity 1 and one group was assigned to Singularity 2, the concepts developed by the groups are described in more detail below.

5.1 Group 1: SQUID

Group 1 looked at the design challenge of AI Personal Assistants. The design question addressed by Group 1 was, “How might we enable the Navy to adapt applications that enhance trust and timeliness in information flow?” The concept created by Group 1 was called the Symbiotic Query Universal Iterative Decision System (SQUIDS).

SQUIDS is envisioned at its core as a question and answer system that allows Naval personnel to get access to experts and to share expertise. In use it is envisioned that SQUIDS will start simple, by connecting experts to people with questions in very specific technical areas, but that it will grow to a broad range of topical areas and that it will be able to provide general support to its users across that range of topics. SQUIDS is intended to be a human-centric tool that supports, not supplants the human by freeing up human cognitive load to hopefully allow humans to focus on creativity, decision making, and other tasks where the team felt humans could provide the most value.

Describing the architecture of SQUIDS, each user will have a profile that describes their interests, expertise, ideas, and the questions they’ve asked and answered in the past. That user’s interface with the overall system is through their personal assistant, called VIBRIO. The user might access VIBRIO through a desktop computer, through a phone or smart watch, or even through an augmented reality display. Each person’s VIBRIO will be connected together through the base SQUIDS AI and database. Access controls will provide the ability to ensure data breaches don’t occur and to protect personal information that might be associated with individual VIBRIO units. The back end database for the global system will consist of a number of different data sources.

Artificial intelligence is assumed to be a necessary part of SQUIDS at several points in the architecture. Each VIBRIO will have an AI that will be expected to learn the preferences of the user it supports. The AI supporting the global SQUIDS back end will be responsible for making connections and for building the database supporting those connections. Because there is an AI involved there is an oversight capability that is needed, to ensure that the AI enhances the human and to ensure that SQUIDS does not limit creativity or have counterproductive behaviors introduced and reinforced. Some amount of this oversight will be provided by the users through their VIBRIO interfaces, some will be oversight of the entire system, and some will be provided by the AI or additional AI.

5.2 Group 2: ADAPT

Group 2 looked at the design challenge of AI Decision Aides. The design question that Group 2 used was, "How might we create an integrated environment that allows decision makers to understand the range of consequences that flow from their decision so they can make informed decisions?".

The concept that was created is called ADAPT, or the Augmented Decision Analysis and Planning Tool. At its core ADAPT is a combination of modeling and simulation (M&S), human machine interface, and course of action (mission plan option) creation. The tool is focused on aiding the deliberate planning process, but could be used for more reactive planning as well. The intended users of the tool are policy makers, COCOM leadership, and planning staff members. The tool will allow the planner to create a course of action in the tool, then will assess that course of action using M&S and artificial intelligence/machine learning techniques. That analysis will take into account intelligence, historical operational performance of both red and blue, and blue force information such as assets available and readiness. Taking the results of the analysis ADAPT will show the planner a range of consequences associated with the plan. The range of consequences are representations of the possible outcomes from the plan and probabilities associated with those outcomes. ADAPT will also recommend potential changes to the initial plan and show the benefits and uncertainties associated with those changes. The planner can use the planning interface to make changes and re-assess the plan multiple times until they get a plan they feel is adequate. At that point the plan can be passed, using the ADAPT tool, to lower level commanders who will take the higher level plan guidance and use it to develop and assess their plans. These more detailed plans will factor into the lower and higher level assessment results. Once the plans are created they will be updated and re-evaluated on a regular basis as new intelligence and commander's guidance is received.

5.3 Group 3: Mind the Gap

This group started with the design challenge: Treating the U.S. Navy as a Complex Adaptive System. After examining the problem, the group reframed the design challenge to "Treating the Navy as a Complex Adaptive Anticipatory Social System". The concept created by the group was called "Mind the Gap". The group described the present situation as one where the capability returns that accrue in the realm of information technology are exponential, whereas the returns that accrue in the realm of human/social/organizational development are linear. This sets up a gap between the two that increases over time, as shown below in Fig. 1.

The desired end state is one where we develop ways to "mind the gap" by improving our organizational ability to deal with complexity. The group's working

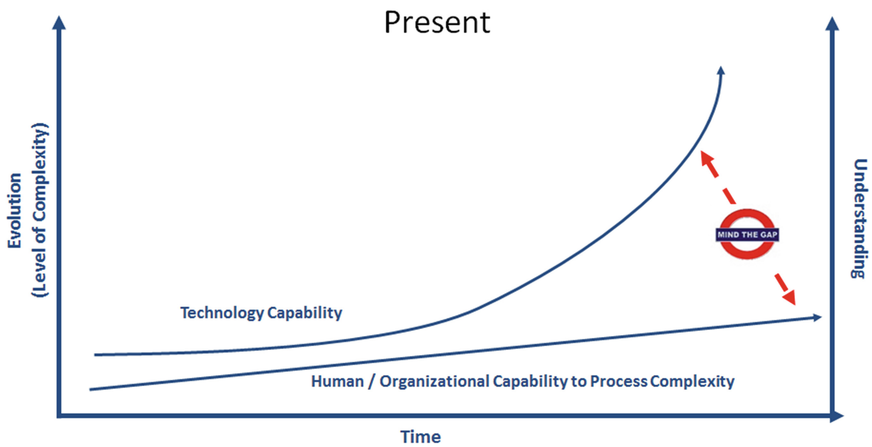


Fig. 1. Gap between technology and organizational capability growth over time

hypothesis was that organizations could meet the challenge of increasing complexity by artfully teaming with technology, as shown in Fig. 2. In this sense, the group's thought process mirrored that of the workshop groups working on Singularity 1, the difference being that the Human-Machine Teaming occurs at the organizational level instead of the individual level.

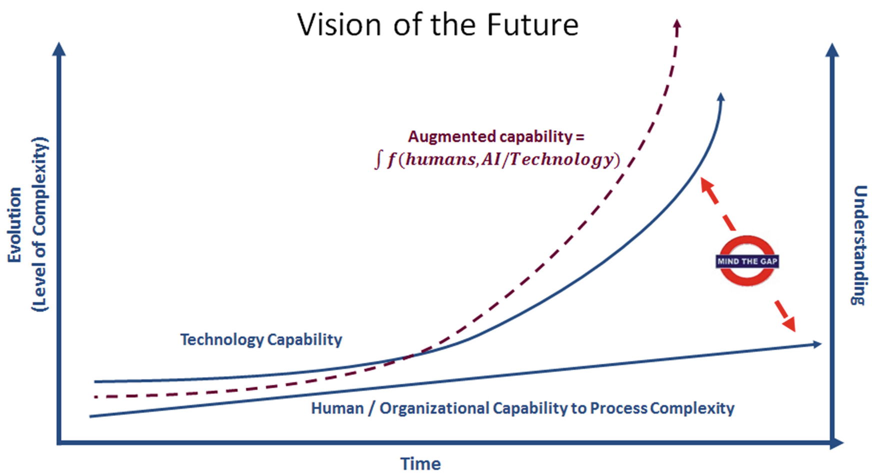


Fig. 2. Impact of human/technology teaming on capability growth

The group developed a number of specific recommendations for closing the gap. These recommendations address two distinct fields. The first set addresses culture at the macro level. The second set consists of specific, tangible studies, projects or pilots which, if undertaken, would produce organizational learning.

Macro Level, Cultural Recommendations:

- Articulate and disseminate across the Navy the Singularity 2 concept - what it is, and why it is important.
- Articulate and disseminate across the Navy a deeper understanding of Tacit Knowledge - what it is, why it is important, and how it fits into organizational capability in the face of rising complexity.

Specific, Tangible Projects, Prototypes:

- Drawing on currently available AI technology, and using recent studies on administrative distractions for targets of opportunity, conduct an exploratory design study, ending in a minimum viable prototype, aimed at using AI to reduce the administrative burden on sailors.
- Drawing on currently available AI technology, conduct an exploratory design study, ending in a minimum viable prototype, aimed at using AI to augment the use of humans in policy enforcement, particularly focusing on contracting and acquisition processes.
- Conduct an exploratory study to assess whether, and if so, how, the diffusion of tacit knowledge might be scaled up by the use of AI.

6 Conclusions

The mmowgli Design for Maritime Singularity game event and following workshop were a unique and creative way to explore the concept of two future Singularities. Through working collaboratively, game players envisioned a future where humans and AI or machines could work together in a complementary manner, thereby increasing the capability of the Human/AI team. Players also envisioned a future where technology such as human/AI teaming combined with changes in organizational construct allowed the U.S. Navy to cope with and even thrive in an increasingly complex environment. The themes identified in the game and the concepts generated in the follow-on workshop can be used by the U.S. Navy and other military organizations to identify future research and development programs to improve upon human/AI teaming and organizational carrying capacity.

Anyone interested in the results of the game can access all of the cards and action plans through the archived reports page [9].

References

1. Schwartz, P.: The art of the long view: planning for the future in an uncertain world. Currency Doubleday Business (1991, 1996)
2. Jensen, G., Tester, J.: Government for the 100%, Using Games to Democratize Innovation and Innovate Democracy. http://www.iftf.org/fileadmin/user_upload/downloads/mmowgli_Government_SR-1539.pdf. Accessed 14 Apr 2018

3. Senge, P.M.: The Fifth Discipline: the Art and Practice of the Learning Organization. Doubleday/Currency, New York (1990)
4. A Design for Maintaining Maritime Superiority. http://www.navy.mil/cno/docs/cno_stg.pdf. Accessed 14 Apr 2018
5. Maritime Singularity Call to Action. https://youtu.be/Oc2zV6hffsY. Accessed 14 Apr 2018
6. Kurzweil, R.: The Singularity Is Near: When Humans Transcend Biology. Penguin (2005)
7. Kasparov, G.: The Chess Master and the Computer. New York Review of Books, vol. 52, no. 2 (2010) http://www.nybooks.com/articles/2010/02/11/the-chess-master-and-the-computer/. Accessed 14 Apr 2018
8. Bar-Yam, Y.: General features of complex systems. Encyclopedia of Life Support Systems (EOLSS). UNESCO, EOLSS Publishers, Oxford (2002)
9. Maritime Singularity Reports page. http://movesinstitute.org/~jmbailey/mmowgliReports/singularity/. Accessed 14 Apr 2018

Step by Step to Peace in Syria

Raphael Parens and Yaneer Bar Yam(✉)

New England Complex Systems Institute, 210 Broadway Suite 101,
Cambridge, MA 02139, USA
yaneer@necsi.edu

The revolution and Civil War in Syria has led to substantial death and suffering, a massive refugee crisis, and growth of ISIS extremism and its terror attacks globally. Conflict between disparate groups is ongoing. Here we propose that interventions should be pursued to stop specific local conflicts, creating safe zones, that can be expanded gradually and serve as examples for achieving a comprehensive solution for safety, peace and stable local governance in Syria.

Nearly five years since the deterioration of government control and order, stemming from the "Arab Spring" protests of 2011, Syria has become a "failed state" with widespread violence, the source of ISIS ideological violent extremism and terrorism, and the origin of massive population displacement and exodus. Often called a civil war, violence in Syria includes conflict between many parties: the government forces, ISIS, several distinct rebel groups with various affiliations, and many external parties including national actors (Saudi Arabia, Turkey, Iran, Russia and the US) and non-state actors (Hezbollah, Druze and Kurdish groups). Despite the existence of alliances on the ground, specifically in opposition to ISIS, these groups have different agendas and support different local groups and may interfere with, or even fight and kill, members of other groups.

Because of the complexity of the conflict there are various framings of the situation as a battle between the opposition and Assad, a battle against ISIS, a battle of proxies, etc. Here we consider a strategy of achieving peace by addressing local conflicts with local solutions prior to building a comprehensive solution to the entire set of conflicts at the national scale. This is a grass roots approach to building safety and security from the ground up. Such efforts already exist in some places. We provide a framework for this approach in terms of a validated scientific analysis of essential drivers of conflict rooted in ethnic geography. While some of those involved will consider our discussion narrowly in terms of whether we provide support for their cause, our objective is to address directly the suffering of populations through establishing robust local safety and governance that will be an immediate relief to local populations severely affected by the conflict. The safety of the populace should not and need not be a hostage for the national solution which may follow.

Complex ethnic geography is a central reason for the large number of different groups in conflict as this geography leads to local allegiances that do not aggregate at the national scale. The makeup, and even existence of several of these groups is a matter of debate, complicated by the possibility that some groups choose to hide their differences (e.g. Alawites and Nusaries) [1]. These debates do not affect the overall framework of our discussion, although they may play a role in the

A. J. Morales et al. (Eds.): ICCS 2018, SPCOM, pp. 380–385, 2018.
https://doi.org/10.1007/978-3-319-96661-8_39

subsequent consideration of local conditions. A well researched map of ethnic geography is shown in Fig. 1 showing 13 ethnic groups that have local geographic areas in which they are the majority population. The largest of these ethnic groups are, in descending order by population size, Arab Sunni Muslims (the majority, 60% of the population), Arabic speaking Alawites, Arabic speaking Christians (Levantines), Kurds, Druze, Ismailis, Nusaries, and Imamis [2].

We have previously demonstrated that, where ethnic groups exist in geographic patches of 20–60 km in diameter, there is a high probability of conflict [3]. We have also shown that establishing local autonomy through subnational boundaries (as is found in Switzerland) is a means of alleviating conflict [4]. Our analysis suggests that providing some level of local autonomy to the ethnic groups would reduce the impetus for local conflict and could serve as a basis for peace and stability [5].

The objective of local autonomy is to enable local decision making to reflect values in determining ordinances, institutions and services to the local population. As is found in Switzerland, where there are multiple cantons of each of catholic and protestant dominated areas, there is no need to have all of the members of a group in a single governance structure. The objective rather, is to reduce friction that leads to conflict by enabling those with widely different values to impose their values on public spaces. We note that the establishment of local partially autonomous regions does not itself determine the structure of federal national governance, which would be the framework for national security and interactions among the local groups as in, for example, the federal government of Switzerland. Other countries have federal governance systems with various ways of balancing local and national governance roles, including the United States, where four levels of governance are present in municipal, county, state and federal systems.

In order to advance the understanding of the conflict and potential solutions in Syria, we have constructed a visualization identifying areas susceptible to violence. This study used a simplified methodology compared to previous studies [3,4]. We superimposed circles of 20 Km diameter on the ethnic map of Syria by visual inspection. Circles are placed where one ethnic population is the majority within the circle and others are the majority around the periphery. Figure 1 manifests why and where ethnic group affiliation, whether defined by language, religion, or both, plays a major role in the conflict in Syria.

Our analytic results for the locations of likely violence are consistent with the current state of conflict shown in Fig. 2. We see that ethnic violence in Syria can be divided into four main regions and an isolated patch: northeast, north, west, southwest, and Deir az-Zur.

Northeast—The northeast region is dominated by Kurd and Sunni Arab groups. Violence is particularly likely between them along the northern border with Turkey, where the two groups are interspersed in 20 Km patches. In addition, some combination of Armenians, Assyrians/Syriacs/Chaldians, and Chechens are surrounded by Kurds and Sunnis in several patches in this area, perhaps also contributing to violence.

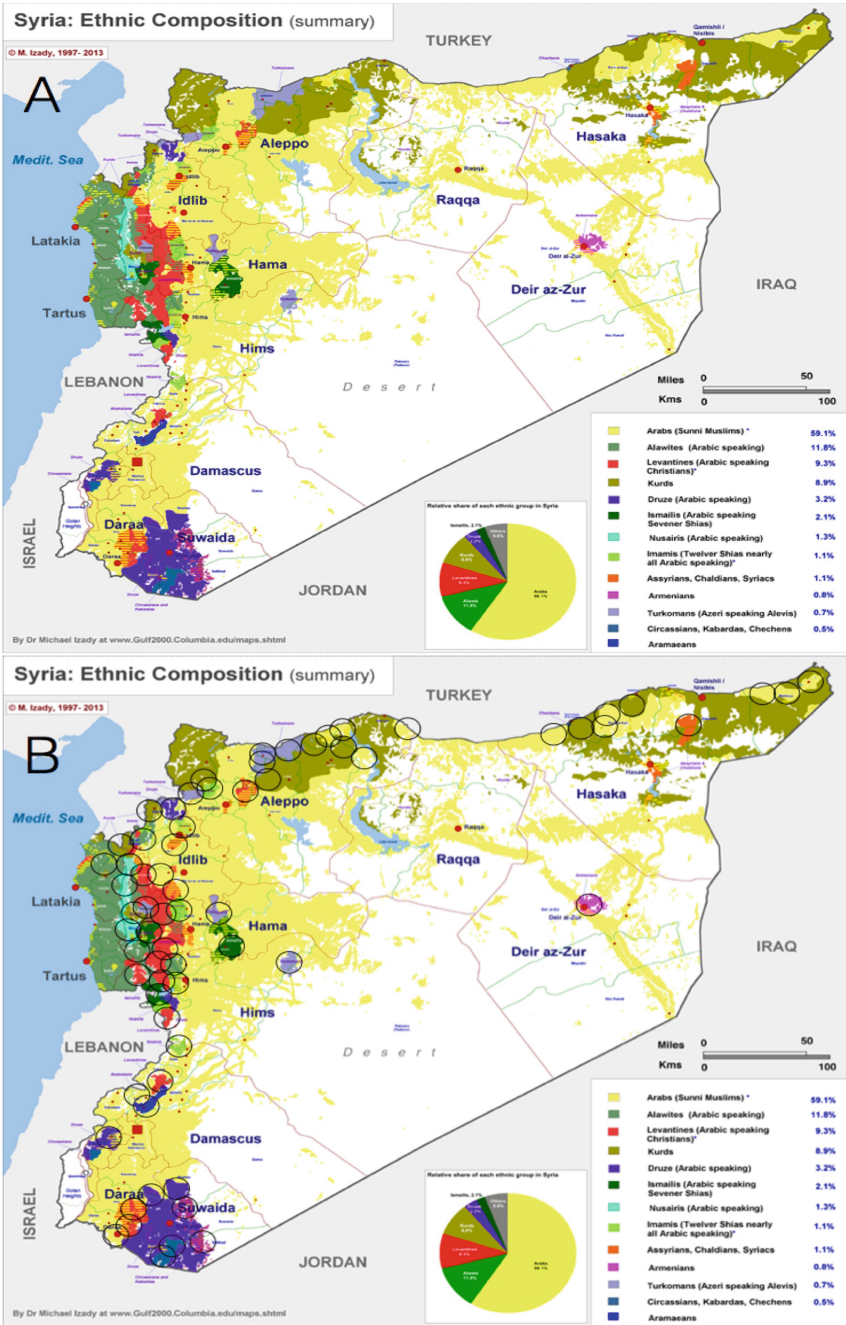


Fig. 1. Ethnic Geography of Syria A. Ethnic geography of Syria [2]. B. Superimposed circles indicate areas of a patch size indicating the likelihood of ethnic violence [3,5].

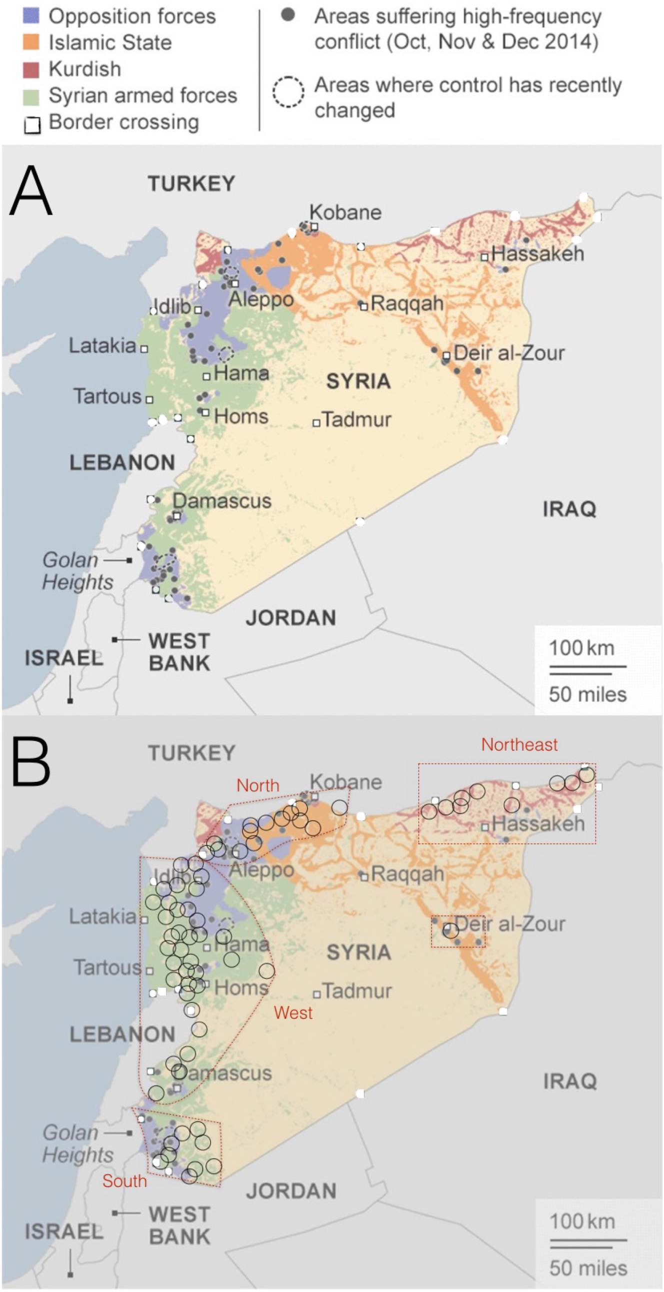


Fig. 2. Conflict in Syria A. BBC map of violence in Syria, July 2015 [6]. B. Superimposed locations of likely locations of violence according to our analysis.

North—The north region contains areas of likely conflict between Kurds, Sunnis, Circassians, Imamis, Levantines, and Turkomans.

West—The west region contains the largest number of likely ethnic violence areas, as a collection of Kurds, Druze, Imamis, Circassians, Assyrians/Chaldians/Syriacs, Ismailis, Nusairis, Turkomans, and Arabs all live in ethnic pockets of the critical size.

South—The south region in the area of Damascus also contains a significant number of ethnic patches, with possible conflict between Levantines, Druze, Aramaeans, Assyrians/Chaldians/Syriacs, Circassians, and Kabardas.

Deir az-Zur—Finally, the city of Deir az-Zur contains an isolated ethnic patch, where a population of Armenians is surrounded by Arab populations.

According to our previous analysis, the establishment of boundaries between ethnic populations to provide partial local autonomy would increase stability and inhibit current and future conflict in Syria. Some suitable natural barriers already exist due to the topography of mountains, lakes, and rivers. However, many boundaries will have to be established through political borders or artificial barriers.

Under the current conditions of fragmentary control and multiple competing groups, implementing a national process to resolve all conflicts is difficult. The complexity of local conflicts, and the many parties involved, will be a barrier to any such comprehensive plan. We propose that a step-by-step bottom up strategy provides a useful and robust alternative to a national plan. Indeed, in this context, the natural scale of intervention is at the community level. Our analysis should serve as a motivation for local governance creation and maintenance in Syria, rather than a blueprint. The complexity of governance creation on the ground will require adaptation due to the specifics of local conditions.

In such a step-by-step strategy "safe zones" should be established. Efforts should be made to identify specific local areas of conflict for intervention to establish the safe zones, including local governance and sub-national political borders, potentially redrawing governorates (muhafazat), districts (manatiq), or subdistricts. Such interventions should recognize specific conditions of villages and urban neighborhoods, their values and capacity to provide for their own security, as well as the relationships they have with nearby groups and national or international groups. The complexity of these local conditions must be addressed through direct engagement with the local population as an integral part of the process of achieving security. In recent years new local governance structures have emerged in response to the disorder [7,8]. These may serve as a basis for the robust and longer-term structures that are needed. Once established, a safe zone will allow the local population to rebuild and reestablish normal lives.

Among the challenges to be faced in achieving safe zones is establishing reliable international support for security, and where necessary, assigning ownership of geographically associated economic resources and religious/cultural site control. Identifying locations where the framework for peace, even if complex, can be more readily achieved is key to early progress and establishing precedent for later more difficult areas.

The greatest challenge in implementing this approach may be the psychological shift from seeing power as an absolute quality on a nation-state basis, and shifting to a perspective in which local power is balanced against the role of allegiances and against larger scales of power, including national. The current conflict is often seen as an irreconcilable power struggle between local groups and the nation-state, as embodied by its government. Here the objective is to show how local groups can coexist with national power. This coexistence is not about full autonomy but rather about a balance between local autonomy within relationships among the local, national and international groups.

By focusing on the local nature of interactions, the basis of community governance can be established. The structure of national power, in whatever form then occurs, becomes secondary as individual day-to-day existence can be primarily determined by the local authority and only secondarily with the national governance.

Acknowledgements. We would like to thank Francisco Prieto-Castrillo, Nima Dehghani and Matthew Hardcastle for helpful comments on the manuscript and Nadim Shehadi, Nassim Taleb and Tom Garvey for helpful discussions.

References

1. Lewis, M.: Michael Izady's amazingly detailed map of ethnicity in Syria (and the Syrian Armenians). GeoCurrents, 26 October 2014. http://www.geocurrents.info/cultural-geography/michael-izadys-amazingly-detailed-map-ethnicity-syria-syrian-armenians#ixzz3yZOkKh7K
2. Izady, M.: Maps and statistics collections. The Gulf/2000 Project (2000). http://gulf2000.columbia.edu/maps.shtml
3. Lim, M., Metzler, R., Bar-Yam, Y.: Global pattern formation and ethnic/cultural violence. Science **317**, 1540–1544 (2007)
4. Rutherford, A., Harmon, D., Werfel, J., Gard-Murray, A.S., Bar-Yam, S., Gros, A., Xulvi-Brunet, R., Bar-Yam, Y.: Good fences: the importance of setting boundaries for peaceful coexistence. PLoS One **9**, e95660 (2014)
5. Bar-Yam, Y., Friedman, C.: Swiss-ification: Syria's best hope for peace. NECSI Report 2013-09-02, 4 September 2013
6. Syria: Mapping the conflict. BBC, 10 July 2015. http://www.bbc.com/news/world-middle-east-22798391
7. Kirkpatrick, D.D.: Fighting shortages, Syrian civilians take reins in rebel areas. The New York Times, 28 February 2013. http://www.nytimes.com/2013/03/01/world/middleeast/syrian-civilians-take-reins-in-test-of-self-government.html?_r=0
8. Hof, F.C.: The self-government revolution that's happening under the radar in Syria. Washington Post, 26 July 2015. https://www.washingtonpost.com/opinions/the-self-government-revolution-thats-happening-under-the-radar-in-syria/2015/07/'6/05cffade-313e-11e5-8353-1215475949f4_story.html

Urban School Leadership and Adaptive Change: The "Rabbit Hole" of Continuous Emergence

Patrick McQuillan[1(✉)] and Brad Kershner[2]

[1] Boston College, Chestnut Hill, MA 02467, USA
mcquilpa@bc.edu
[2] Carolina Friends School, Charlotte, NC 28273, USA

Abstract. In the current educational context deliberate and continuous emergence seems eminently logical. Schools comprise so many interacting dimensions—moving parts of people, ideas, contexts, and resources—change truly is the norm. School systems therefore need to adjust to both the challenges and opportunities they regularly encounter. In doing so, the principal represents a critical leverage point. Believing that "Leadership is no longer the activity of gatekeeping and directing but of enabling and empowering" (Morrison 2002, p. 19), wrote that administrative leaders should "enhance the skills and knowledge of people in the organization [and] create a common culture of expectations around the use of those skills and knowledge" (p. 15). One strategy for addressing this challenge and enacting these ideals is to generate a complex adaptive system in which power and authority are decentralized and all school personnel—students, teachers, administrators, and parents—embrace a common vision committed to shared beliefs, values, policies, and practices. Accordingly, we conceptualize systems emergence as an adaptive process in which a school "system" adjusts to its context, drawing upon the analytic heuristic known as *continuous emergence* to reveal the ongoing and intertwining challenges that arise for urban school leadership when this occurs. In terms of the emergence process, we engage the experience of *disequilibrium*, *intensification*, *emergent order*, and *stabilizing feedback* not as linear phenomena leading to a single outcome but as an ongoing process in which these features of emergence interact in ways that are largely non-linear and unpredictable yet still reveal promising strategies for adapting to the varied sources of disequilibrium that arise in the system.

Keywords: Complexity theory · Complex adaptive systems · Emergence

[I]nstead of privileging equilibrium and equilibrating change, organization science should treat disequilibrium and disequilibriating change as natural and ongoing rather than exceptional and episodic. … Equilibrium need no longer be viewed as the natural state to which a system returns (Chiles et al., p. 514).

A. J. Morales et al. (Eds.): ICCS 2018, SPCOM, pp. 386–397, 2018.
https://doi.org/10.1007/978-3-319-96661-8_40

1 The Challenge of Urban School Leadership

In the current educational context deliberate and continuous systems emergence seems eminently logical. Schools comprise so many interacting dimensions—moving parts of people, ideas, contexts, and resources—change truly is the norm. Standardized exams, fundraising committees, instructional leadership teams, lock-down drills, and a host of other practices and related expectations were largely non-existent 15 years ago. For urban principals, these developments are even more pronounced, as they serve disproportionate numbers of low-income, geographically-mobile, special education, immigrant, and non-native-English-speaking students (Hemmings 2012). To date, the impact of these challenges has been pronounced. Consider these longstanding trends: An achievement gap on the National Assessment of Educational Progress that "doggedly persists for Latino/Latina and Black students, notwithstanding billions of dollars spent yearly on remedial and compensatory intervention programs" (Valencia 2015, p. 8); a national drop out rate, concentrated in cities, that currently exceeds three-quarters of a million students annually; and turnover rates among urban teachers, principals, and superintendents that undermine efforts to generate any sense of school community or consistent student achievement (Payne 2008). By many measures, urban schools struggle, and so do their principals.

Given this context, urban school systems would benefit from becoming adaptive, able to adjust to both the challenges and opportunities they regularly encounter. In doing so, the principal represents a critical leverage point in a school system that comprises myriad interacting elements and dimensions (Fullan 2005; Davis and Sumara 2006), embracing a role characterized as "enabling and empowering" (Morrison 2002, p. 19) while seeking to "enhance the skills and knowledge of people in the organization [to] create a common culture of expectations around the use of those skills and knowledge" (Elmore 1996, p. 15). To do so schools need to become more adaptive, to generate a complex adaptive system in which power and authority are decentralized and all school personnel—students, teachers, administrators, and parents—embrace a common vision committed to shared beliefs, values, policies, and practices.

To understand the nature of these demands and the relevance of this adaptive strategy, we offer a case study of Richard Davidson and the Jeffrey Jackson School[1] as a means to conceptualize emergence as an adaptive process, one in which a school "system" adjusts to its context in ways that promote equitable outcomes for all involved. To conceptualize the nature of this challenge, we draw upon the analytic heuristic known as *continuous emergence* to reveal the ongoing and intertwining challenges that arise for urban school leadership, engaging the experience of *disequilibrium*, *intensification*, *emergent order*, and *stabilizing feedback* not as linear phenomena that constitute emergence leading to a single outcome but an ongoing process in which these four features of emergence interact in ways that are largely

[1] All names are pseudonyms.

unpredictable yet still reveal promising strategies for successfully adapting to change. Accordingly, we see complex change as potentially both a *deliberate* and *continuous process*. As MacIntosh and MacLean (1999) observed:

> [T]he key difference in applying complexity theory ... to organizations as opposed to organisms is that organizations have the capacity to bring about a change in archetype through consciously creating the conditions in which successful transformation can occur. Those within the organization can, to some extent, choose the primary rules which govern the deep structure (p. 306).

also contrasted organisms and organizations to illuminate the potential of the tenets of complexity theory to promote deliberate emergence:

> Like natural living systems, generative human systems are also in a state of continuous dynamic balance. But the presence of human intention—purpose and meaning—adds conscious choice to the system's dynamics, influencing the nature and flow of its information, the diversity and transparency of its relationships, and the complexity of its processes, patterns, and structures (p. 31).

2 Initial Conditions at Jeffrey Jackson School

The Jeffrey Jackson School (JJS), a K-5 elementary school in Boston Public Schools, is not the prettiest Boston public school but its outward appearance belies a sense of community and care that permeates much of school life. Classrooms and hallway bulletin boards are colorful and full of positive messages in English and Spanish: "You are entering a learning zone"; "Parents are our partners." Historically, JJS has been seen as a solid public school where students received a quality education. In 2013 JJS became one of four high-performing, high-poverty schools in BPS. JJS achieved the highest evaluation, Level 1 status, between 2013 and 2016, the end of this study, even though the school enrolled 84% English language learning students, 23% of whom were illiterate in Spanish. In 2015, only nine students didn't pass MCAS, the Massachusetts standardized exam.

Richard Davidson was principal at JJS from Fall 2011 to June of 2016, the period encapsulated in this study. Deeply committed to ensuring all students succeed, during his tenure he worked to bring students, teachers, and parents into a coalition that worked to continually enhance student achievement. When he became principal of JJS, he was the school's third principal in three years. He succeeded a principal with an authoritarian style and intended to change JJS leadership dynamics to a more inclusive, decentralized structure. Early in his tenure, Richard reflected on this situation:

> The principal before me was very effective at leading with a top-down authoritarian style.... Although results were evident, the leadership method was not in full accordance to my mind-set, skill, and conviction.... During many initial conversations [they seemed to] need direction, decision making and guidance. I often asked, "How was this done last year?" The staff consistently echoed, "The principal made all decisions."

His first year at JJS, Richard sought to establish his presence as a supportive, trustworthy leader. As one teacher remarked:

> He really works hard to make the situation best for the individual [teacher]. If it means he has to cover someone else's class so that they can go work with that particular child or that particular group, that's what he'll do.... It's the first time I've seen that.

Teachers also appreciated that Richard trusted them, whereas previous principals encroached on their autonomy. In one teacher's words:

> [I was] micromanaged to the point that I didn't want to do [extra work] anymore. [Richard] has not done that at all. He's pretty much handed it over and said, "You know what you need to do. Keep me posted if there are any issues." ...I just feel like he's trusting me to get done what has to get done.... It's put less stress on me because I know what needs to get done.

Although faculty liked Richard, being third in line in a rapid succession of school leaders hindered his ability to transmit a sense of urgency to teachers—leaving him in a difficult position. In Richard's first year a veteran teacher reflected on the impact of rapid turnover:

> [W]e've had three different principals in the four years I've been here. And each principal has their own missions and their own ideas of what they want the school to become.... [F]or the staff, it's kind of hard to relate to each [principal], not knowing how long they're going to be here or how much time they should put into any new initiative if someone else is going to come in and change it.

A younger teacher outlined Richard's predicament in a way that seemed both cynical and revealing:

> To make [change] happen, there is a toll that has to be paid. Before you get on the good side of these people—who are very, very, very set in their ways—you have to sort of placate them first. No matter what your values are, if they're at all different from theirs, you have to placate them.... You have to be harmless and helpful. If you are those two things only then will any of those teachers care at all what you think about anything.

Although complexity science emphasizes non-linear, and therefore non-predictable, outcomes, within complex adaptive systems forces known as "attractors" lend a measure of predictability to how systems operate by shaping behavioral patterns (Wheatley 1999). In schools, culture—what people believe and value—can influence behavior notably. At JJS, Richard regularly told faculty "We are crew, not passengers," aiming to promote a shared commitment to a vision of successfully educating every student.

Overall, Richard brought a keen commitment to ensuring student achievement to his work at JJS. Some teachers embraced this ideal while others questioned Richard's motivations and professional competence.

3 The Rabbit Hole

In our view, the current situation facing urban schools, such as JJS, demands constant adaptation. Myriad pressures never allow schools to remain in a state of equilibrium for long. Dissonance and disruptions are omnipresent. But this is not necessarily bad.

New ideas and promising directions constantly emerge, promoting intensification and perhaps a new emergent order, one more effective than those previous.

Fully embracing this reality, Richard Davidson aimed to create change at JJS. His efforts were informed by three guiding principles derived from his work at a Boston College-based leadership academy: to share power and authority with faculty; to make faculty active participants in the school's efforts at instructional leadership; and to promote a common school vision embraced by all members of the JJS community. The enactment of these ideals did not occur in a uniform fashion. Richard often turned to particular faculty to serve on key committees and to share ideas linked to various reforms. To understand the outcomes of Richard's efforts we parallel the four steps characteristic of complex emergence (MacIntosh and Maclean 1999)—*disequilibrium*, *intensification*, *emergent order*, and *stabilizing feedback*—as a means to conceptualize emergence. In contrast to others, we describe this process as a continuous phenomenon not a discrete, linear experience, and this has implications for how emergence as institutional change is enacted.

To begin, somehow, in the context of the multiple factors that impact complex change, the routine must be disrupted. Systems do not change if elements within the system continue doing what they have always done. For adaptive change to emerge, some sense of *disequilibrium* (Nadler 1993; Wheatley 1999)—be it external "turbulence" originating from the broader context or internal "perturbations" intentionally generated to promote adaptive change (Beabout 2012)—must create a state where "the system is ripe for transformation…. experiencing new opportunities, new challenges, and new ways to understand the world" (Reigeluth 2004, p. 27). As Wheatley (1999) observed, disturbing the status quo is critical to transformation:

> [A]nything that disturbs the system plays a crucial role in helping it self-organize into a new form of order…. a higher level of complexity, a new form of itself that can deal better with the present. In this way …*disorder* can be a source of new *order*, and …growth appears from disequilibrium (p. 21; emphasis in original).

When conceptualizing emergence, a 2nd dimension, *intensification,* represents a state when processes within institutions gain momentum, often because of mutually reinforcing features of system interactions. Power is often distributed more broadly, creating increasingly decentralized networks where new structures emerge as heretofore nonexistent connections are generated. New ideas and practices surface. Some are challenged, some endorsed, some are transformed. A sociocultural vision continues to emerge as foundational beliefs, values, practices, and processes interact to promote systems adaptation, iteratively and recursively. As the processes unfold, actors have opportunities to reveal their competence, sincerity and reliability, setting the possibility to enhance relational trust and lay a foundation for transformation (Bryk and Schneider 2002; Moolenar and Sleegers 2010). In this context, effective leaders drive conversations and collaborative efforts, promoting their sociocultural vision whenever the opportunity arises, engaging with faculty, parents, and students in iterations of change around a sociocultural vision.

The third dimension to emergence entails *emergent order*, the state when synergistic interactions among system elements generate a fluid structure that can adapt to its contextual challenges and environment. It is something new and transformed, capable of further evolution and adaptations. It has become a complex adaptive system (Stacey 1996) which, through a recombination of resources, increases the capacity of the overall system to operate in a state of temporary equilibrium as agents and resources in the overall system re-combine into new patterns of interaction that tend to improve system functioning by adapting to the demands emerging in the overarching context.

The fourth and final dimension of the emergence process, *stabilizing feedback*, entails modifications applied to the initial stages of an innovation, manifest as technical adjustments not adaptive transformation (Lichtenstein and Plowman 2009). While emergent order entails transformation of an existing system where something new has evolved, the system may still be dysfunctional, needing adjustments to realize its transformative potential, encouraging positive developments and discouraging those that undermine system performance. This final stage of emergence is not about system transformation, *per se*; it entails minor modifications to the status quo so the system continues doing what it is now doing but more efficiently and effectively. MacIntosh and MacLean (1999) spoke to this issue: “In the final stage of the conditioned emergence model, feedback is applied to amplify actions consistent with the new rules and archetype. Feedback must also be used to damp actions or behaviors which belong to the old rules” (p. 311).

To understand aspects of what occurred at JJS under Richard’s leadership this paper draws upon the notion of continuous emergence, viewing emergence as an intertwined and ongoing process of often unpredictable but overlapping demands on the institution and its actors. To provide a sense for how this dynamic played out at Jeffrey Jackson School we present two emails Richard Davidson sent to one author, Patrick McQuillan.

4 Richard Davidson Emails: April 8, 2014 and October 15, 2015

Two emails from Richard Davidson on April 8, 2014 and October 15, 2015—the first being two-and-a-half years into his JJS tenure and the second slightly more than four years—are revealing, as both manifest matters central to the emergence process: disequilibrium, intensification, emergent order, and stabilizing feedback, all arising in a continuous and notably unpredictable process. In the first missive Richard wrote:

> Hello Pat,
> I have been 100% consumed with the multiple, concurrent, and overlapping initiatives that have different levels of district prioritization. I have been trying to participate in the dual advisory boards I have been appointed to, conduct family engagement activities, crunch the budget, facilitate data cycles, interview [job] candidates, enter interview data, set up the PARCC [a new state exam], set up technology to support the PARCC, [teacher] evaluations, evaluation input, cut staff, have difficult conversations with those staff, deal with the staff who have now “circled the wagons” around the staff who have been cut, facilitate PD [professional development], get staff weaned off Reading Street [a reading program the district no longer supported], organize interventions, implement interventions, and deal with an aging staff that has looming retirements this year and next. I am working to keep

the ship moving in the right direction despite veteran staff having a hard time letting go and transitioning into retirement…. As a sole administrator, who has delegated some leadership aspects of the school, life is difficult right now.

How do I manage six to seven different entities for the school while still being an instructional leader? After having a family engagement piece, a fundraising piece, a partnership-building piece, a marketing piece, an instructional learning piece, and a lot of the other stuff in the human resource side. That is everything. How does one manage all that and still keep up with all the emails?

With that being said, I am helping my oldest son get into college and my youngest son transition into middle school, and finish my dissertation [Richard had enrolled in a doctoral program].

Ms. Arnette and Veronica Jackson [two administrative staff] are maxxed out as well…. Needless to say there is no let up.

Regards,

Richard Davidson (April 8, 2014)

A year-and-a-half later, McQuillan received this email from Richard, again touching on all dimensions of the emergence process:

Hello Pat,

I have been rolling along since the second week in August when I reported for work. Tomorrow [Saturday] I have our annual Fall Festival from 12 to 3. Next week, I have our award presentation from the Office of Student Engagement regarding our earning of the "Family Friendly Certification." October 31st is Parent University. I am working to get a solid core of parents there. I have decided to provide a bus but charge parents and I am not providing childcare at school. I am trying to see if this affects family attendance to Parent U…. I have had two great staff professional development sessions. September 3rd I was at [a local site] for an all-staff professional development. Last week I had a staff PD that was differentiated and addressed a couple instructional issues. I have developed and launched the following leadership teams: Instructional Leadership Team, Positive Behavior Intervention Team, Math Leadership Team, Book Room Development Team, Student Support Team, Operations Team and an ESL/ Reading Intervention team. I have had to give speeches regarding being the recipient of an Innovation Award from [a local educational foundation]. I had a visit the third day of school with the [BPS Superintendent], [a representative] from the Mayor's office, and a few others. I am working to get everyone on board in the construction of a shared instructional focus and shared vision for this year and the school. We as a school are trying to embrace the district's mantra of "A culture of WE!"

The Operations Team is one of our premier teams that is working to become effective and efficient problem solvers. I have setup a rotation for afterschool coverage so we all share the burden. We are currently reading the book, *The Five Dysfunctions of a Team*…. Our next Instructional Leadership Team is Wednesday 10/21/15 at 8:00 a.m.

Needless to say, there is a lot going on. . .

Richard Davidson (October 15, 2015)

Both emails allude to all four stages in the emergence process, but certainly not as a uniform process. At first (4/8/14), *disequilibrium* emerged from "the multiple, concurrent, and overlapping initiatives" that generated a milieu of uncertainty and dissonance, in part, because they all had "different levels of district prioritization." Serving on dual advisory boards created opportunities for Richard to gain new insights and impress his supervisors but they demanded time and attention, thus requiring him to balance new opportunities, often associated with disequilibrium, with ongoing demands on his time. Interviewing job candidates always represents an opportunity tension and related disequilibrium because one never knows what impact those persons may have but new hires inevitably have some impact (Hramiec 2017). Finally, Richard had to set up a computer system for the PARCC exam in a school with limited technological capabilities. In a telling statement, Richard noted, "As a sole administrator, who has delegated some leadership aspects of the school, life is difficult." In only delegating "some" leadership responsibilities, relational trust seemed an issue (Bryk and Schneider 2002). Without trusting more people to assume responsibilities, Richard's workload and related disequilibrium remained substantial.

In similar fashion, a year-and-a-half later significant uncertainty characterized aspects of what occurred at JJS, simply in terms of what would happen with newly created committees—the Positive Behavior Intervention Team, Operations Team, and Reading Intervention Team, being potential sources of disequilibrium. Further, on the third day of school representatives from the Mayor's office and superintendent's office would visit the school. In both instances, the sources of disequilibrium seemed notable, blending promising opportunities with challenging responsibilities.

The two emails also suggested aspects of system *intensification*. In the first, disaffected faculty "circled the wagons" around staff Richard had dismissed, requiring him to reaffirm and rearticulate his commitment to a shared sociocultural vision. In the later missive, Richard alluded to "working to get everyone on board in constructing a shared instructional focus and shared vision," including "the district's mantra of, 'A culture of WE!'" Further, faculty were reading *The Five Dysfunctions of a Team*. These actions provided teachers with opportunities to dialogue about new ideas and new ways to conceptualize their professional responsibilities. These also offered Richard opportunities to highlight the values, beliefs, and practices he considered central to the work of JJS faculty and administration.

Both emails also embodied emergent order, signs that some system elements were enacting the sociocultural vision in a coherent, mutually reinforcing manner, creating a state of temporary equilibrium and adaptability, a complex adaptive system. The first mentioned "family engagement activities," data cycles, and "facilitating PD [professional development]," interrelated elements of the JJS system with synergistic potential linked to enriched teaching and learning driven by positive interactions with parents and teacher professional inquiry. In the later communication Richard described "two great PD sessions" that were "differentiated" and linked to "instructional issues," suggesting that mechanisms were in place to promote instructional leadership and distribute authority as means to enrich teacher commitment and professional expertise. There was also an annual Fall Festival, and upcoming Parent University and Family Friendly Certification award presentation, each signaling significant JJS-parent interaction. The school's sociocultural vision also seemed to have taken hold in terms of some faculty seeing themselves as crew-not-passengers, as Richard noted seven recently created teams, all intertwined emergence with instructional leadership, thereby offering opportunities to reinforce the common sociocultural vision. Working in concert, these elements of the somewhat transformed JJS status quo represented an adaptive system, possessing features that facilitated information flow, a shared vision, and personal autonomy. Productive routines were in place to help a school "system" adapt, endure, and in some instances, thrive.

Richard also mentioned a tension that suggested concern with according all faculty power and authority, as some work felt overwhelming: "How do I manage six to seven different entities for the school while still being an instructional leader?" Not all pieces were in place for the system to thrive. Some faculty seemed to lack authority and commitment to shared ideals and practices. Systems features were in place but not all faculty were part of systems; rather, an overall assessment would suggest that new networks had come into existence but they were to some degree "fragmented" (Davis et al. 2012). Yet Richard also noted that the Operations Team was "working to become effective and efficient problem solvers," that the school had setup a rotation for afterschool coverage so "We all share the burden," and the ILT will meet again soon, all signs that power and authority had been shared and was being enacted, at least somewhat.

Richard's notes also alluded to stabilizing feedback, adjustments that enhanced interactions in some newly formed networks without transforming the system. In the first email this included "crunch[ing] the budget," a means to allocate funds where Richard felt they were most needed, a responsibility which could enrich specific aspects of the JJS school "system" but would be unlikely to generate system transformation. Likewise, Richard's opportunities to evaluate teachers' practice offered opportunities to influence their practice and reinforce a shared sociocultural vision. In the second note Richard alluded to the Operations Team's effort to create an equitable plan for after-school coverage in which "all share the burden." Further, Richard tweaked aspects of his work with parents, maintaining involvement with Parent University but lowering expenses and requiring less support from JJS faculty. All of these actions aimed to enrich systems interactions but not transform systems. They targeted "technical" not "adaptive" challenges, matters that could be resolved by expertise and attention.

As emergence unfolded as a continuous process new networks formed which themselves continued to evolve and intertwine. In so doing, one can see all four dimensions of the emergence process in the work being undertaken at JJS.

5 Conclusions

In general, Richard's efforts to promote change at JJS seemed relatively successful. As the emails suggest, Richard's commitment to sharing power and authority with teachers, to enact instructional leadership, and promote a common school vision have led to the creation of something like a complex adaptive system that can endure and grow in an urban school context but he did not have the full support from all faculty.

Comments from a reading specialist at the end of Richard's tenure highlight how JJS embarked on the process of change as a continuous phenomenon:

> It's been a good and exciting year, and I think that we've had the freedom this year, because a lot of other things that were previous challenges are less so, so that I can really dig really deep on pedagogy and the academic work, and we've come to a lot of interesting conclusions, a lot of interesting questions, and it's sort of like that rabbit hole goes deeper and deeper and deeper (May 24, 2016).

Given the work undertaken at JJS and the implications linked to continuous emergence derived from the two emails, the following insights linked to systems change seem relevant:

> *Challenges are opportunities.* Disequilibrium often arises as challenges emerge. Yet challenges also represent opportunities to reconceptualize one's work and generate more satisfying and effective outcomes. At JJS, Richard noted, "As a sole administrator, who has delegated some leadership aspects of the school, life is difficult." In a related vein, in the second note Richard alluded to the creation of new committees. In both cases disequilibrium for Richard and others emerged from the uncertainty and, to some measure, a lack of relational trust between Richard and JJS faculty, pointing to two critical issues: Should Richard trust faculty with additional authority and how much is appropriate? And when he did, what would they do with this new-found power?

In seeking to address those challenges, evolve solutions don't mandate them (Lewin 1999). This means going slower but doing so allows you to bring more people into the process of change, thereby generating dialogue and enhancing relational trust—key elements to successful change (Bryk and Schneider 2002). In the work at JJS, this would be central to the process of Richard relinquishing power to faculty and their consequent efforts to enact that power. The first step would be to relinquish some power, monitor the consequences of such efforts, and then determine appropriate next steps.

In the process of evolving change, link change to a shared cultural vision. As a system, the elements that make up school systems interrelate. Certainly, a school's cultural vision represents a critical aspect of the overall system, which will intertwine with and emerge at multiple system levels, if the system is operating effectively. You therefore will encounter many opportunities in varied systems contexts to draw out the relevance of your shared vision. Don't pass on any opportunity to do so because shared values lay a foundation for subsequent change and for enacting those changes in ways that maintain fidelity to the overarching vision. Summarizing the notion of "principal as culture builder," nearly 50 years ago Sarason (1971) wrote:

> Life for everyone in a school is determined by ideas and values, and if these are not under constant discussion and surveillance, the comforts of ritual replace the conflict and excitement involved in growing and changing.... If the principal is not constantly confronting one's self and others, and if others cannot confront the principal with the world of competing ideas and values shaping life in a school, he or she is an educational administrator and not an educational leader (p. 177).

Finally, not all changes are created equal. Looking at the various manifestations of emergence that arose at JJS reveals notable differences. For the purposes of understanding systems change it is useful to note that some matters were rather technical, setting up a computer system for the PARCC exam, for instance. Time and technical expertise could address that concern. In contrast, faculty PD and newly created leadership teams represented efforts where people enacted power in new and different ways, and would therefore benefit if guided by a shared sociocultural vision, one in which practices, policies, beliefs, and values aligned to promote outcomes linked to that vision. This would mean that attention would need to be directed toward promoting specific values, practices and beliefs that would encourage actions on the part of JJS faculty and staff that aligned with this vision.

References

Beabout, B.: Turbulence, perturbance, and educational change. Complicity **9**(2), 15–29 (2012)

Brown, S.L., Eisenhardt, K.M.: The art of continuous change: linking complexity theory and time-paced evolution in relentlessly shifting organizations. Adm. Sci. Q. **42**(1), 1–34 (1997)

Bryk, A.S., Schneider, B.: Trust in schools: a core resource for improvement. Russell Sage Foundation, New York (2002)

Chiles, T.H., Meyer, A.D., Hench, T.J.: Organizational emergence organization science. Organ. Sci. **15**(5), 499–519 (2004)

Davis, B., Sumara, D.J.: Complexity and Education: Inquiries into Learning, Teaching, and Research. Routledge, New York (2006)

Davis, B., Sumara, D.J., D'Amour, L.: Understanding school districts as learning systems: some lessons from three cases of complex transformation. J. Educ. Change **13**, 373–399 (2012)

Elmore, R.: Getting to scale with good educational practice. Harv. Educ. Rev. **66**(1), 1–25 (1996)

Fullan, M.: Leadership and Sustainability: System Thinkers in Action. Corwin Press, Thousand Oaks (2005)

Hargreaves, A., Fullan, M.: Professional Capital: Transforming Teaching in Every School. Teachers College Press, New York (2012)

Hemmings, A.: Urban High Schools: Foundations and Possibilities. Routledge, New York (2012)

Hramiec, A.: Hiring: The Very First Step to a Flourishing School Culture. Students at the Center Hub (2017). http://studentsatthecenterhub.org/blog/hiring-the-very-first-step-to-a-flourishing-school-culture-part-3-of-3/

Kershner, B., McQuillan, P.J.: Complex adaptive schools: educational leadership and school change. Complicity Int. J. Complexity Educ. **13**(1), 4–29 (2016)

Lewin, R.: Complexity: Life at the Edge of Chaos. University of Chicago Press, Chicago (1999)

Lichtenstein, B.B., Plowman, D.A.: The leadership of emergence: a complex systems leadership theory of emergence at successive organizational levels. Leadersh. Q. (2009). https://doi.org/10.1016/j.leaqua.2009.04.006

MacIntosh, R., MacLean, D.: Conditioned emergence: a dissipative structures approach to transformation. Strateg. Manag. J. **20**(3), 297–316 (1999)

Moolenaar, N.M., Sleegers, P.J.C.: Social networks, trust, and innovation: the role of relationships in supporting an innovative climate in Dutch schools. In: Daly, A.J. (ed.) Social Network Theory and Educational Change, pp. 97–114. Harvard Education Press, Cambridge (2010)

Morrison, K.: School Leadership and Complexity Theory. Routledge, New York (2002)

Nadler, R.: Therapeutic process of change. In: Glass, M. (ed.) Adventure Therapy: Therapeutic Applications of Adventure Programming, pp. 57–69. Kendall/Hunt Publishing, Dubuque (1993)

Payne, C.: So Much Reform, So Little Change: The Persistence of Failure in Urban Schools. Harvard Education Press, Cambridge (2008)

Reigeluth, C.M.: Chaos theory and the sciences of complexity: foundations for transforming education. Paper presented at the annual meetings of the American Educational Research Association, San Diego (2004)
Sarason, S.: The Culture of the School and the Problem of Change. Allyn & Bacon, Boston (1971)
Stacey, R.D.: Complexity and Creativity in Organizations. Berrett-Koehler Publishers, San Francisco (1996)
Valencia, R.: Students of Color and the Achievement Gap: Systemic Challenges, Systemic Transformations. Routledge, New York (2015)
Wheatley, M.J.: Leadership and the New Science: Discovering Order in a Chaotic World. Berrett-Koehler Publishers, San Francisco (1999)

Reading the Media's Mind

Sarjoun Doumit(✉) and Ali A. Minai

Department of Electrical Engineering & Computer Science,
University of Cincinnati, Cincinnati, OH 45221-0030, USA
doumitss@mail.uc.edu, ali.minai@uc.edu

Abstract. 'There's no art to find the mind's construction in the face,' wrote Shakespeare, but trying to infer what someone is really thinking is arguably the essence of interaction between cognitive agents. Turning this into a computational model is challenging, but one possible approach to infer mental models from linguistic expression is to look at patterns of lexical associations. Assuming that written language reflects conceptual associations in the writer's mind, we have previously shown differences in the patterns of lexical association between creative and non-creative writing. In this paper, we apply the same approach to news reports from individual media sources over the same period, with the goal of looking for differential associative patterns. The underlying assumption is that the associative patterns of a media source will reflect its "mind" and "personality," i.e., specific styles, preferences, or biases, just as they do for individuals.

Keywords: Cognitive · Graph theory · Machine learning

1 Introduction

How is news reported? The answer often is that it depends on which media reported the news, despite it being the same piece of news. This is partly because the topic of the news story itself largely determines where people go to learn about the underlying events and the path they take to get there [8]. Typically, people do not simply rely on a few primary sources for their news, but tend to follow a suite of media feeds (includes RSS, podcasts etc.) – often with a similar inclination.

The study of the news content can reveal interesting characteristics of any media's mindset and allow them to be compared to each other. Analysis of news content has been a central focus for media scholars, political scientists, sociologists and historians. The availability of on-line news makes it possible to analyze it on a large scale using methods developed in the fields of Web intelligence, data mining and machine learning. The issues that news content analysis tries to address include identification of salient topics, summarization of stories, extraction of opinions, construction of semantic networks, and characterization of news reports in terms of bias. These are also the motivating issues for our research.

A. J. Morales et al. (Eds.): ICCS 2018, SPCOM, pp. 398–405, 2018.
https://doi.org/10.1007/978-3-319-96661-8_41

This paper describes a framework which consists of two systems that we have previously developed: IONA [10] and ALAN [4,5,7,9]. IONA (*Intelligent On-line News Analysis*) uses an iterative form of *latent Dirichlet allocation* (LDA) for the analysis of online corpora with large numbers of small news documents, extracting the summary of news stories from RSS feeds and capturing the potential biases of media sources. ALAN (*Analysis of Lexical Association Networks*) is a statistical framework for building lexical association networks from text corpora and analyzing them in terms of their network properties. We have previously applied this approach to analyzing the writings of poets and authors to assess their associative styles. Here, we consider different media sources to be *cognitive agents*, and analyze the corpora of their reports to characterize their "mindsets" in a comparative fashion.

The work we describe has two main goals: (1) Comparing the network properties and reporting preferences of different media sources for the same time period, (2) Comparing the network properties and reporting preferences of the same media source at different time periods, especially where similar events have happened in the time periods. This allows us to observe the evolution of potential biases in a single media source's mindset on the same topic over time, and also to characterize how different media sources choose to report the same news.

2 Related Work

Currently, several research and commercial systems are available for analyzing and clustering textual news in order to gain insight from it. The methods underlying these systems range from purely statistical to graphical models. These models typically provide information that must be interpreted and organized further by a human user. For example, WEIS [11] and CAMEO [6] are both systems that use *event analysis*, i.e. they rely on expert-generated dictionaries of terms with associated weights, and parse the text to match the words from the news event to those in the dictionary. They can then categorize the information into a set of expert-defined categories with respect to sentiment intensity values. Other systems, such as Oasys2.0 [3] use a construct called *opinion analysis*, which depends on user feedback rather than on experts in order to determine the intensity value of an opinion. The Oasys2.0 approach is based on aggregation of individual positive and negative references identified [1].

3 Methods

We have been collecting and building an extensive database covering many online world-wide news media sources through their English-version RSS feeds to test our analysis approach. We collect all news articles from these media sources round-the-clock at specific time intervals and store the data organized by media source. The news is aggregated at the end of every month and stored in a database as shown in Fig. 1.

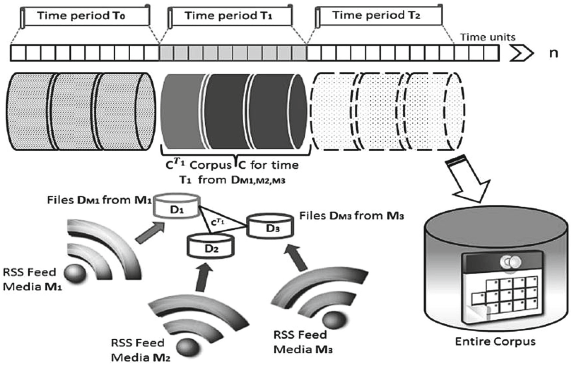


Fig. 1. RSS news harvesting in IONA.

IONA uses LDA, an unsupervised generative algorithm, to discover latent topics in text corpora [2]. It is based on the postulate that every document is a mixture of latent topics, where each topic is a multinomial distribution over the vocabulary of the corpus. In its original form, LDA requires the user to choose the desired number of latent topics in the corpus. IONA's improvement on this restriction is a heuristic that uses a feedback-loop that builds *lexical motifs* from the output of LDA, and combines them into evolving *motif chains* until a threshold is reached. That heuristic results in: (1) Creating semantically coherent *super-motifs* that self-organize into clusters representing relevant latent stories, and (2) Tagging of the corpus' documents using the latent stories that were discovered in step 1. ALAN is a tool for extracting and processing text from different sources, constructing several types of *lexical association networks* (LANs), and computing statistical and network metrics for these. Several measures of word association can be used in ALAN, including joint (co-occurrence) probability, correlation coefficient, log odds ratio, and pointwise mutual information (PMI), all defined relative to RSS story as the unit [4,5,7,9]. Network metrics generated include mean shortest path length, degree, clustering coefficient, and various centralities – calculated at both the node and network level.

The combined framework first constructs *story-graphs* from a news corpus using IONA. The documents from which these stories came are tagged as relevant and the rest of the reports are removed. This filtered corpus of reports is then fed into ALAN, which computes associations between words in all the stories. These associations are then mapped back to the original story graphs to label the associations in each story with a strength value that can be used to analyze the news source's reporting preferences and biases.

4 Experiments

To test our results, we chose an event that has occurred twice in recent years: Reported chemical weapons attacks on civilians allegedly by the Syrian government in August 2013 and April 2018. The news regarding both chemical attacks was covered by all world media and we chose to test this event in our framework to see what we can discern, especially regarding the aftermath of each event. The media sources we chose were the BBC, CBS, France 24, Jerusalem Post and Reuters. We ran IONA for these two one month periods and picked up all the main stories during these two time periods – including both events in Syria. The important point is that, for each reporting period, all media sources were confronting the same world events, so the differences and similarities in their reporting – which words they emphasized or ignored, which words they associated strongly or weakly, etc. – can be used to analyze how each news source chose to place and report their story.

Figures 2 and 3 show the reports on the 2013 Ghouta event by the Jerusalem Post and the BBC respectively. Each graph is a selection of subgraphs from the story graph repertoire of the respective source from this period. Subgraphs that used at least one term associated with the Syria story were included. The weight on an edge from word j to i indicates the conditional probability of i occurring in a window of text that contains j. These probabilities are supplied by ALAN. Words emphasized by a double border indicate that they were among the top ten words (by frequency) for that media source in the reporting period.

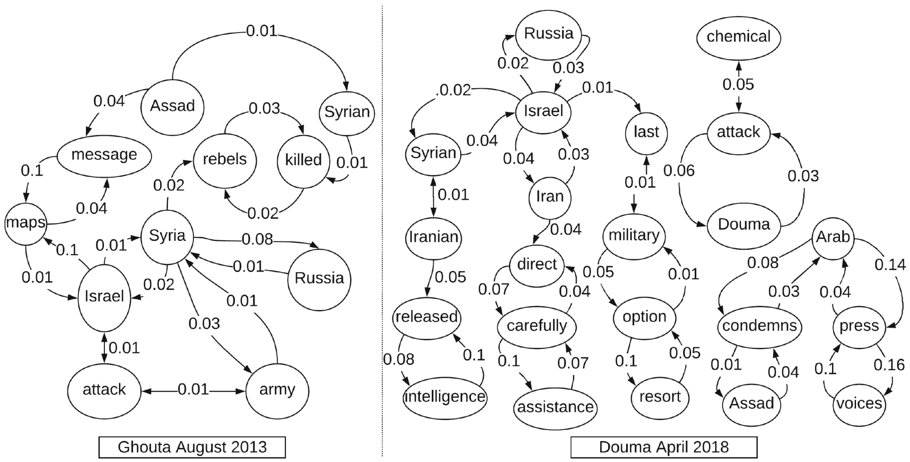


Fig. 2. The Jerusalem Post's coverage of the chemical attacks in Syria for 2013 and 2018

Our first analysis shows that the Jerusalem Post's telling of the Ghouta story is much more *vigorous* than that of the BBC, in a way that connects many more words to the overall story in addition to spawning smaller independent stories that are also relevant (not shown). The BBC's Ghouta story shared elements with the Yemen story and was generally vague in comparison to the Jerusalem Post. But that changes for the next analysis when we look at the their coverage for the later Douma attack: the story-graphs interestingly are much more

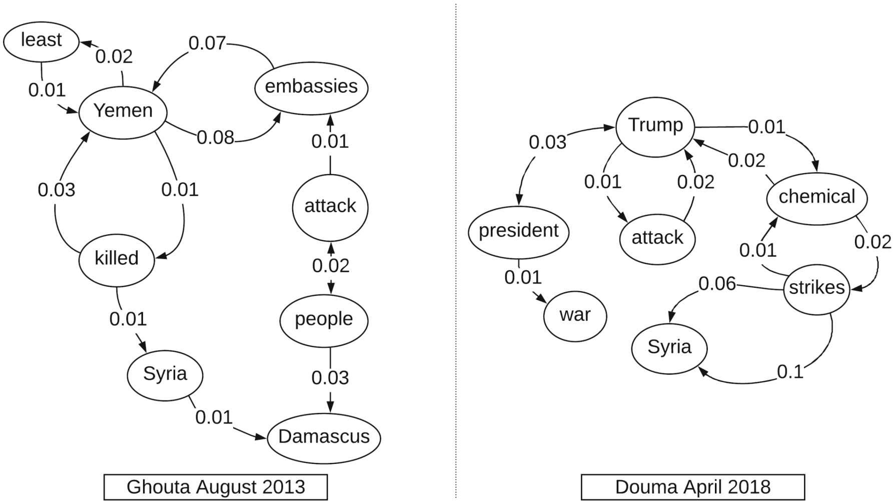


Fig. 3. The BBC's coverage of the chemical attacks in Syria for 2013 and 2018

coherent with the main story agglomerating more words in a more cohesive network. This is evident from words such as *Iran* being pulled into the overall Syria chemical attack story. It is interesting to note that the Douma story reported by the BBC is centered significantly on President Trump – presumably because his response to the attack was a major concern in the UK.

To assess how much significance each source gave to the stories, we calculate two metrics:

1. The *Relative Word Count* is the ratio of the number of unique words associated with the story divided by the total number of unique words in all stories from that source for the period.
2. The *Relative Story Weight* is the ratio of the summed frequencies of unique words associated with the story divided by the summed frequencies of unique words in all stories from that source for the period.

Looking at these metrics for both attacks shown in Table 1, we can see that CBS has increased its *interest* in this topic, Reuters increased slightly, while the other media sources maintained their significance levels.

Looking at the network metrics for significant words using the joint probability for the BBC (Table 2) and for the Jerusalem Post (Table 3) we see that the numbers tell of a much larger overall shift for BBC than the other news media, and a small change for Jerusalem Post. This suggests that, while the reporting style of the BBC seems to have evolved, those of the other media have remained stable. An interesting observation is that the name of the town in the first chemical attack (Ghouta) did not even show up in the metrics in the 2013 report, but 5 years later, both town names involved in the attacks were present, showing that news reports were referencing the previous attack and therefore

Table 1. Analysis of the alleged Syrian chemical attack in Ghouta (2013) and Douma (2018) individual *story* from the overall *stories* for all 5 media sources

		Syria chemical attacks story graphs analysis	
		Relative word count	Relative story weight
2013	BBC	0.0677	0.11
	CBS	0.0617	0.19
	France 24	0.0141	0.0630
	Jerusalem Post	0.071	0.274
	Reuters	0.017	0.077
2018	BBC	0.0574	0.095
	CBS	0.1272	0.6998
	France 24	0.0317	0.0544
	Jerusalem Post	0.06	0.2346
	Reuters	0.142	0.142

adding the older story to the narrative. It should be noted that this analysis uses only RSS feeds, which are very brief on-line reports. The name of 'Ghouta' was used widely in longer media reports, but they are not part of our dataset.

Table 2. Network metrics from the Joint Probability associations of the BBC for the chemical attacks in Syria stories

Joint	**BBC August 2013**					**BBC April 2018**				
Probability	Degree	Eigenvector Centrality	Betweeness Centrality	Closeness Centrality	Clustering Coefficient	Degree	Eigenvector Centrality	Betweeness Centrality	Closeness Centrality	Clustering Coefficient
Syria	196	0.006625	24213.29	0.000398	0.111773	268	0.001926	102298.56	0.000123	0.072484
Chemical	154	0.008978	8710.68	0.000387	0.1553147	195	0.001604	40192.33	0.00012	0.105154
Attack	312	0.00951	62128.77	0.000438	0.0750474	383	0.002798	232399.14	0.000128	0.60875
Douma						73	0.000764	6577.52	0.000109	0.208904
Ghouta						22	0.00022	366.29	9.4E-05	0.411255
Rebels	46	0.002308	815.32	0.000346	0.3594202	38	0.000371	1485.37	0.0001	0.291607
Assad	35	0.001873	220.69	0.000324	0.4638655	31	0.000367	622.74	9.9E-05	0.417204
Iran	59	0.001495	2916.10	0.000346	0.2466393	44	0.0003	3171.07	9.9E-05	0.18816
Israel	36	0.001177	1229.13	0.000328	0.3539682	44	0.000312	4118.10	9.9E-05	0.163847
Russia	20	0.00125	49.96	0.000311	0.7	121	0.000879	26929.36	0.000111	0.112745

Finally we show the top 10 words from BBC, Jerusalem Post and France 24 for the covered period (Table 4). These are the top words across all stories, not just the Syria attack stories, and the comparison is telling. All three

Table 3. Network metrics from the Joint Probability associations of the Jerusalem Post for the chemical attacks in Syria stories

Joint Probability	Jerusalem Post August 2013					Jerusalem Post April 2018				
	Degree	Eigenvector Centrality	Betweeness Centrality	Closeness Centrality	Clustering Coefficient	Degree	Eigenvector Centrality	Betweeness Centrality	Closeness Centrality	Clustering Coefficient
Syria	250	0.010007	69462.51	0.000479	0.065196	267	0.012999	105850.81	0.000468	0.049618
Chemical	159	0.008978	18378.56	0.000456	0.129368	95	0.005752	10923.58	0.000388	0.140873
Attack	185	0.009979	26289.36	0.000463	0.113506	99	0.006815	15376.43	0.000403	0.142238
Douma						54	0.003927	3175.90	0.000362	0.252271
Ghouta						16	0.001199	109.54	0.000326	0.833333
Rebels	49	0.002515	2212.05	0.00039	0.264184	70	0.003983	14250.27	0.000376	0.174741
Assad	212	0.01013	43785.33	0.00047	0.089054	97	0.00574	14047.98	0.000392	0.127147
Iran	65	0.003396	4498.38	0.000393	0.192788	109	0.005429	28126.97	0.000406	0.102457
Israel	100	0.004554	17977.55	0.000431	0.126984	155	0.009252	39808.35	0.000428	0.096439
Russia	34	0.002149	904.27	0.000374	0.385026	72	0.005759	8222.76	0.000396	0.178012

Table 4. The top 10 words from BBC, Jerusalem Post and France 24 for both time periods

Freq.	BBC		Jerusalem Post		France 24	
	August 2013	April 2018	August 2013	April 2018	August 2013	April 2018
1	News	Games	Syria	Syria	France	President
2	World	Common wealth	Syrian	Iran	French	French
3	Day	Women	Assad	Trump	Syria	France
4	Killed	World	Attack	Israel	President	Trump
5	Attack	Trump	Report	Syrian	Country	Killed
6	New	First	Leader	Iranian	Government	Macron
7	Hours	Win	Chemical	Chemical	Egypt	Syria
8	People	President	Strike	Attack	New	New
9	August	New	Response	Strikes	Friday	Strikes
10	Pictures	Syria	Obama	Reported	Killed	State

sources mention the attacks, but the news is clearly of less significance for the BBC, somewhat greater significance for France 24, and of major concern for the Jerusalem Post – which is natural, considering the geography and history of the sources. It is also interesting to note that Iran appears as a significant word in the April 2018 Jerusalem Post reports, but not in those of the BBC and France 24. And, as the story graph for the Jerusalem Post story shows, that source links Iran to the attack – something that is not seen in the other two sources.

5 Conclusion

We have demonstrated an interesting and promising framework that can potentially be used for the comparative analysis of different media to understand their styles, preferences, emphases, and biases. Almost all of the process is fully automated, which means that it can be used for real-time analysis of many media sources on a large scale.

Acknowledgment. This work was supported in part by National Science Foundation INSPIRE grant BCS-1247971 to Ali Minai.

References

1. Benamara, F., Cesarano, C., Picariello, A., Reforgiato, D., Subrahmanian, V.: Sentiment analysis: adjectives and adverbs are better than adjectives alone. In: International AAAI Conference on Weblogs and Social Media (ICWSM), pp. 203–206 (2007)
2. Blei, D., Ng, A., Jordan, M., Lafferty, J.: Latent Dirichlet allocation. J. Mach. Learn. Res. **3**, 993–1022 (2003)
3. Cesarano, C., Picariello, A., Reforgiato, D., Subrahmanian, V.: The OASYS 2.0 opinion analysis system. In: International AAAI Conference on Weblogs and Social Media (ICWSM), pp. 313–314 (2007)
4. Doumit, S., Marupaka, N., Minai, A.A.: Thinking in prose and poetry: a semantic neural model. In: Proceedings of IJCNN (2013)
5. Doumit, S., Minai, A.A.: Effect of associative rules on the dynamics of conceptual combination in a neurodynamical model. In: Proceedings of IJCNN (2015)
6. Gerner, D., Abu-Jabr, R., Schrodt, P., Yilmaz, O.: Conflict and mediation event observations (cameo): a new event data framework for the analysis of foreign policy interactions. In: International Studies Association of Foreign Policy Interactions (2002)
7. Marupaka, N., Iyer, L.R., Minai, A.A.: Connectivity and thought: the influence of semantic network structure in a neurodynamical model of thinking. Neural Netw. **32**, 147–158 (2012)
8. Media Insight Project & Associated Press-NORC Center for Public Affairs Research: The personal news cycle: how Americans choose to get their news. American Press Institute (2014)
9. Mei, M., Vanarase, A., Minai, A.A.: Chunks of thought: finding salient semantic structures in texts. In: Proceedings of IJCNN (2014)
10. Doumit, S., Minai, A.: Semantic knowledge inference from online news media using an LDA-NLP approach, vol. 21. IEEE (2011)
11. Tomlinson, R.: World event/interaction survey (WEIS) coding manual. Department of Political Science, United States Naval Academy, Annapolis (1993)

Social Neointeraction on Facebook, Presidential Campaign Mexico 2018

Carlos Augusto Jiménez Zarate(✉)

Faculty of Administration and Public Accounting of the Autonomous University of Nuevo Leon (FACPYA-UANL), Pedro de Alba S/N, University City, San Nicolás de los Garza, N.L., México
uanl.jimenez@gmail.com

Abstract. In 2008, there was a turning point in the history of electoral campaigns, the first presidential campaign of Barack Obama was the beginning of policy 2.0. Obama's campaign team demonstrated the power and influence in modern society, of the social-digital networks and diverse platforms such as Google, YouTube, Twitter, instant messaging and websites, included in what is called information technologies and communication (ICT). Ten years after that mythical election, a new paradigm becomes more evident. They are the 4.0 networks which will interconnect almost everything to the internet, within this new dynamic, the social networks are also evolving, today it is not just about uploading or viewing content, but to interact and thereby generate broadcast networks. Within the spectrum of social media, there is one that has become, in the cornerstone of modern communication. Facebook is the most important social network at international level, its evolution has been vertiginous, and it went from being a network among friends, to become an indispensable advertising platform in the world of digital marketing, to such a degree that it has positioned itself as the second medium, recipient of advertising spending, just below Google. Facebook is a multiplatform, where you can find applications, games and pages known as "fanpages", these websites represent, for companies, political parties, public figures and for governmental and civil organizations, a virtual system of social interaction. When users comment on a publication, discussion or debate networks can be generated among users, increasing the organic scope of the publication. The present investigation will focus on analyzing the structure of the networks that generate the likes in the comments made, within the fanpages of the presidential candidates in Mexico, who have positioned themselves as leaders in most of the electoral polls, for the presidency from Mexico. The set of data and information obtained as: degree of interactions, reactions, discussion, debate, support, comments and responses, can be called "neo social interaction". This result will serve to have an alternative vision, about the popularity, from the point of view of the interactions in their fanpages, taking into account that Facebook has positioned itself as the most influential medium for obtaining political and electoral information.

Keywords: Social network analysis · Social neointeraction · Viral reaction
Viral sentiment · Networks on Facebook

A. J. Morales et al. (Eds.): ICCS 2018, SPCOM, pp. 406–413, 2018.
https://doi.org/10.1007/978-3-319-96661-8_42

1 Introduction

The interaction between users of social-digital network platforms such as Facebook, generates complex systems, which are structured in free-scale networks [1]. A set of nodes and links are called network that represent the interaction or interconnection between the elements of the set. Most of the basic studies of social networks focus on determining the centrality, influence and connectivity of the nodes in a network [2]. Social interaction on Facebook can be represented by networks, where users are nodes, and the type of interaction (likes, comments or reactions) among users is shown as links or edges.

If a Facebook user reacts, comments or shares a post, it is likely that their friends can see and interact with the post; this in turn creates new interactions, thus generating organic viral interaction-diffusion networks, which can form cascades of more than two degrees of separation [3]. Sharing a publication is the action of greater engagement with a publication, it means that a user is totally in agreement and takes the publication as its own. The set of interactions, reactions, sharing or comments generated by the publications in the social-digital networks, such as Facebook, is proposed to be called social neointeraction. Facebook's fanpages have become the new mass medium, for marketing and for all areas of society in general. This platform offers various applications and tools, for the management and maximum use, within the area of digital marketing. According to the ZentihMedia ranking, Facebook is the medium that has grown the most in recent years, in terms of advertising revenue, ranking second in the ranking of 2017, second only to Google. Between Google and Facebook, 20% of advertising spending is concentrated globally. In politics, the power of influence of the networks of friends on Facebook, to make the decision to go to vote, is largely due to strong ties, which represent about 7% of friends or contacts of each user [4].

Within social media platforms, Facebook has become a fundamental axis in political campaigns, through electoral marketing [5]. Currently, virtually all candidates have a fanpage on Facebook. Mexico is currently experiencing a historic electoral contest, for the first time a candidate identified with the policies of the left, is leading almost all the polls. This race has been divided into three stages: Pre-campaign (December 14, 2017 to February 11, 2018). Inter-campaign (from February 12 to March 29, 2018) and the Campaign (from March 30 to June 28, 2018). For the coalition between the "Institutional Revolutionary Party" (PRI, by its initials in Spanish), "Green Party" (PVEM, by its initials in Spanish) and the "New Alliance Party" (PANAL, by its initials in Spanish) is headed by José Antonio Meade Kuribreña (MEADE). For the coalition between the National Regeneration Movement Party (MORENA, by its initials in Spanish), the Social Encounter Party (PES, by its initials in Spanish) and the Labor Party (PT, by its initials in Spanish), Andrés Manuel López Obrador (AMLO) is in charge. For the coalition between the National Action Party (PAN, by its initials in Spanish), Citizen Movement (MC, by its initials in Spanish) and the Party of the Democratic Revolution (PRD, by its initials in Spanish), is headed by Ricardo Anaya Cortes (ANAYA). Independent candidates stand out Margarita Zavala (ZAVALA) and Jaime Rodriguez Calderón. (BRONCO) The vote to elect the next president of Mexico will be held on July 1, 2018.

1.1 Purpose

This research aims to obtain the networks generated by candidates for the presidency of Mexico, from their fanpages on Facebook, in the period called pre-campaign. From the obtained networks the structure will be analyzed, from the distribution of the degree of each node, its trend and its equation will be obtained, in addition to the network metrics. In previous studies, network metrics have been correlated with the vote obtained, this in electoral processes for governor in Mexico [6].

This research integrates the quantification and qualification of the likes in the comments in the fanpages of the politicians, who aspire to the presidency of Mexico. The metrics of the structure of the network, the type of interaction and the feeling of the comments, together represent the “social neointeraction” that is generated in a fanpage. This information is a complement to the prospection of popularity and the intention to vote that candidates can reach. Within the network metrics, the centrality, intermediation and influence of the nodes in the network will be analyzed, which are of vital importance, in order to know the topology. Other measures, such as the average shortest path length, are used to observe the efficiency of the transmission of information or published content [7]. The dynamics of social influence have been changing since the 1980s, thanks to greater social interconnection via the internet and social media, the party bias has been decreasing and there has been a growing independence in electoral behavior [8].

2 Neointeraction Networks on Facebook

Internet users who have registered their profile on Facebook, have a wide diversity, to perform actions and interactions between their network of contacts or friends, likewise can be “fans” (followers) of various fanpages. In a Facebook fanpage you can publish a variety of content, text, photos, videos, external links and also applications. When a user comments on these publications, which can be supportive (positive) or against (negative) or neutral. Comments can generate “threads”, since other users or followers can respond to comments and create a conversation or discussion, while other users can react to comments, with a like, or an emoticon. These actions originate networks with free-scale characteristics [9], with a greater power of interaction-diffusion, since the friends of the users who interacted could see these actions reflected in their news feed.

Of the networks generated by this interaction, two equations can be obtained: The first is the network equation, which reflects the distribution of the nodes and their degree. The second equation, represents the viral equation that represents the likes that managed to generate the comments. Social neointeraction networks on Facebook, can reach to form three zones. The first is the zone of direct likes (Reacted post), are those users or followers of the fanpage that only give like to the publications, this type of reaction is located within the first level of interaction-social diffusion, since only some of your friends or contacts will see this action. The second zone, and of greater relevance are the comments, which the users make to the publications, since this action can be seen by more friends or contacts, increasing the diffusion of the publication. The third zone is made up of the reactions through likes or emoticons to said comments, this area is called in this research viral reaction (see Fig. 1).

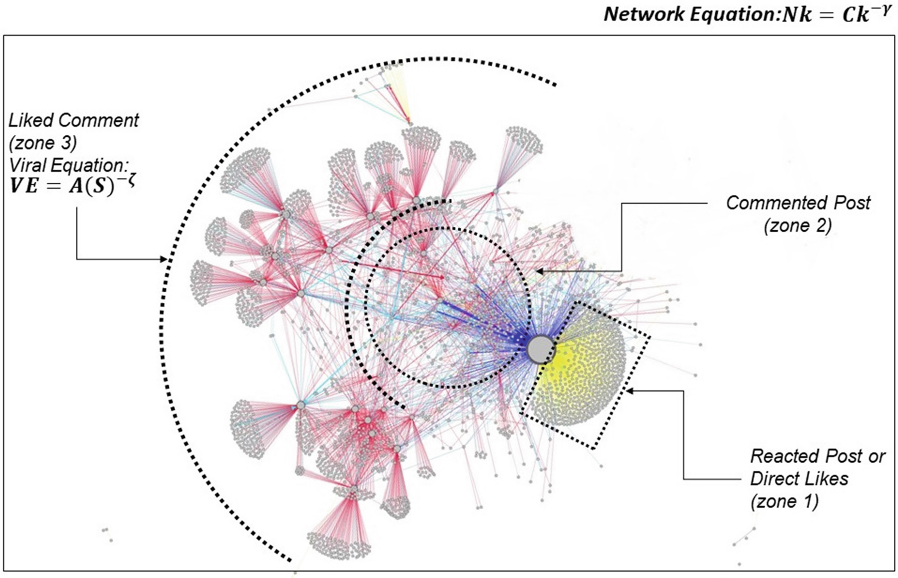


Fig. 1. Social neointeraccion network on Facebook. The different zones of the network are defined, as well as the equations of the structure.

3 Methodology

The NodeXL software was used, to obtain a sample of the interactions between the comments and the likes of each fanpage analyzed, the sampling period was from January 21 to January 23, 2018. The metrics were calculated using the Gephi software. The visualization of the data obtained was done using the YiFanHu algorithm [10]. The network metrics, such as centrality and structure [11, 12] of the analyzed fanpages, can offer greater clarity of the structure and dynamics of viral interaction-diffusion of social networks [13].

The distribution of the degree of the nodes of the sampled networks has the form of the distribution of the power law, a distinctive feature of complex systems, which is represented in this research by the following nomenclature of the network equation:

$$Nk = Ck^{-\gamma} \tag{1}$$

Where:
Nk = Nodes of the network with "k" links.
C = Constant of the equation.
k = Number of links (degree) of the nodes.
$-\gamma$ = Coefficient of the network equation.

To measure the power of the viral reaction, it is proposed to use the following equation:

$$VE = A(S)^{-\zeta} \tag{2}$$

Where:
VE = Equation of the viral reaction (sub-likes).
A = Constant of the equation.
S = Comments with Sub-likes.
$-\zeta$ = Network viral equation coefficient.

For the qualification of the viral sentiment (%), it is proposed to weigh the comments, cataloging into two categories: Positive comments towards the candidate will be given a value of one, and negative or neutral comments a value of zero.

The formula proposed for the qualification of the comments, will be called "viral sentiment" (ςv).

$$\varsigma v = \frac{\sum_{i=1}^{n}\{(\mathrm{L}_i)(S_i[1\, or\, 0])\}}{\sum_{i=1}^{n} L_i}(100) \tag{3}$$

Where:
L_i = Number of sublikes that generated the comments analyzed.
S [1 *or* 0] = Sentiment of the comment analyzed, if it is positive its value will be one (1), if it is negative or neutral it will be zero (0).
n = Number of comments analyzed.

4 Results

The network graphs of each fanpage show clear differences between the structures of social neointeraction developed. ANAYA, MEADE and AMLO present networks with viral reaction zones, but in the AMLO network, the communities have a larger number of users and a structure of more free-scale interconnection (See Fig. 2).

The results of the network metrics show that the AMLO network had a higher average path length, this type of metric can indicate in the case of a fanpage network, the scope of a publication. Another relevant metric was the modularity index, almost all of them were below .5, but not that of AMLO, which achieved a modularity index of .615, this metric indicates the strength of the communities created. In the EigenVector metric, the highest value was obtained by the BRONCO fanpage, which was one of the candidates with the highest negative viral feeling index. While in the clustering metric it was the ANAYA fanpage that reached an index of 0.439, this metric is an indicative of the interconnection force of neighboring nodes. The coefficient of the network equation (power law) is used to define a complex system, this coefficient must have a

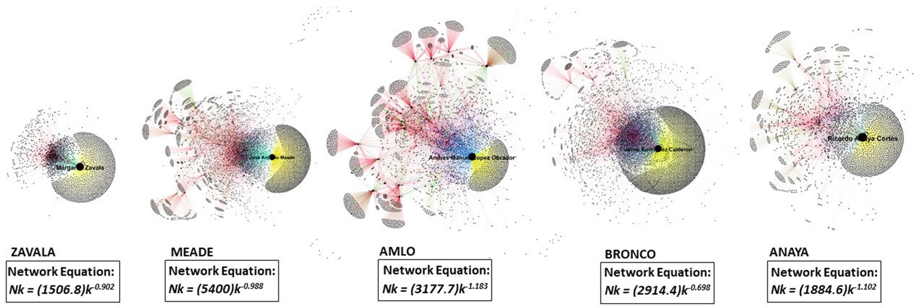


Fig. 2. Fanpages networks on Facebook, of the presidential candidates. Mexico 2018.

value between −2 and −3, the result of this research (Table 1) shows that only the network coefficients, the fanpages of AMLO and ANAYA are greater than −1.

In Table 2, we can see the type of interaction that users did in the fanpages of the candidates for the presidency of Mexico.

The results of the fanpage's viral equations showed that the AMLO fanpage obtained the highest percentage of "liked comment", in addition to its viral equation coefficient greater than −1, with a positive viral sentiment of 95%. AMLO also achieved the highest average path length of the network, this metric is an indication of the viral power of diffusion or interaction, in the modularity metric was the fanpage that obtained a higher index.

Table 1. Network metrics, from the fanpages on Facebook of the presidential candidates. Mexico 2018.

	ZAVALA	MEADE	AMLO	BRONCO	ANAYA
Network equation	$Nk = (1506.8)\,k^{-0.902}$	$Nk = (5400)\,k^{-0.988}$	$Nk = (3177.7)\,k^{-1.183}$	$Nk = (2914.4)\,k^{-0.698}$	$Nk = (1884.6)\,k^{-1.102}$
Nodes	2341	7851	5037	8035	2935
Edges	6264	18823	8733	18365	5207
Avg. Degree	2.676	2.398	1.734	2.286	3.548
Network diameter	13	15	33	24	4
Avg. Path Length	3.636	5.103	9.96	3.288	2.639
Graph density	0.001	0	0	0	0
Modularity	0.115	0.427	0.615	0.307	0.558
Eigenvector centrality	0.04427	0.00678	0.0062	0.11609	0.06505
Avg. Clustering coefficient	0.108	0.071	0.049	0.128	0.439

Table 2. Interaction of the users in the fanpages of the presidential candidates. Mexico 2018.

	ZAVALA	MEADE	AMLO	BRONCO	ANAYA
Liked comment	*38.83%*	*47.90%*	*56.19%*	*34.79%*	*42.85%*
Viral equation (sub-likes)	VE = 223 $(S)^{-0.78}$	VE = 1493 $(S)^{-0.992}$	VE = 2058 $(S)^{-1.265}$	VE = 278 $(S)^{-0.73}$	VE = 385 $(S)^{-0.973}$
Positive viral sentiment	*100%*	*18%*	*95%*	*17%*	*100%*
Negative/Neutral viral sentiment	*0%*	*82%*	*5%*	*83%*	*0%*
Reacted post	25.05%	15.06%	11.18%	22.15%	21.47%
Commented coment	9.56%	10.16%	11.59%	9.90%	15.38%
Consecutive commenter	18.74%	17.22%	11.34%	22.93%	14.19%
Commented post	6.90%	8.79%	9.25%	9.24%	4.90%
Tagged	0.61%	0.68%	0.44%	0.90%	0.63%
Shared	0.32%	0.19%	0.02%	0.09%	0.08%

The viral sentiment measurements of the ZAVALA and ANAYA networks gave 100% positive sentiment results, but these results have a low impact, according to the viral equation of their fanpages.

The fanpage of ZAVALA was the one that obtained more "direct likes" (Reacted Post), this type of reaction has a low degree of social interaction.

5 Discussion and Conclusion

Social networks such as Facebook have become the main source of news in the United States of America; In the case of Mexico, according to the recent study titled "The habits of Mexican Internet users", of the Internet Association, 92% of users, seek information about electoral processes, and 97% of users do so through social networks, among which Facebook stands out first with 95%. This research focused on the social neointeraction networks, which are developed once the users, react to the comments in the publications of the fanpages, of the candidates for the presidency of Mexico. Of the analyzed networks, the one with the highest coefficient, in the distribution of the degree of the nodes, was the candidate of the coalition MORENA-PT-PES, Andrés Manuel López Obrador (AMLO), with a coefficient of the equation of the network $-\gamma = 1.183$. AMLO also obtained the highest viral equation coefficient of $-\zeta = 1{,}265$. The results of both coefficients can give us an overview of the dynamics and social structures that originate in Facebook fanpages. And although both coefficients are not in the characteristic range of complex systems, they can be a reference for the complex networks generated in the sociopolitical dynamics. This research proposes that if a fanpage of some candidate in political campaign, has the following values ($-\gamma > 1$) and ($-\zeta > 1$), it is considered a complex sociopolitical system of interaction-diffusion viral on Facebook. Several electoral polls indicate that López Obrador (AMLO) is the candidate with the greatest intention to vote [14]. It can be concluded according to the data obtained that we, that AMLO, is the candidate that causes the social neointeraction, with the greatest positive viral impact on the Facebook fanpages of the presidential candidates in Mexico.

References

1. Barabási, A.: From network structure. IEEE Control Systems Magazine, 33–42, August 2007. https://doi.org/10.1109/MCS.2007.384127
2. Newman, M.E.J.: The structure and function of complex networks. Siam Rev. **45**(2), 167–256 (2002). https://doi.org/10.1137/S003614450342480
3. Anderson, A., Huttenlocher, D., Kleinberg, J., Leskovec, J., Tiwari, M.: Global diffusion via cascading invitations: structure, growth, and homophily. In: Proceedings of the 24th International Conference on World Wide Web (WWW 2015), pp. 66–76 (2015). https://doi.org/10.1145/2736277.2741672
4. Bond, R.M., Fariss, C.J., Jones, J., Kramer, A.D.I., Marlow, C., Settle, J.E., Fowler, J.H.: A 61-million-person experiment in social influence and political mobilization. Nature **489** (7415), 295–298 (2012). https://doi.org/10.1038/nature11421
5. Postigo, M.: Campaña en la red: estrategias de marketing electoral en Internet. Revista Académica de Marketing (2012). http://dialnet.unirioja.es/servlet/articulo?codigo=4125640&info=resumen&idioma=SPA
6. Jiménez Zarate, C.A.: Dinámica y efectividad de las fanpages de Facebook de candidatos a gobernador en los resultados electorales. Innovaciones de Negocios, **13**(26), 221–238 (2016). http://www.web.facpya.uanl.mx/rev_in/Revistas/13_26/13.26 A4.pdf
7. Boccaletti, S., Latora, V., Moreno, Y., Chavez, M., Hwang, D.: Complex networks: Structure and dynamics, **424**, 175–308 (2006). https://doi.org/10.1016/j.physrep.2005.10.009
8. Braha, D., De Aguiar, M.A.M.: Voting contagion: Modeling and analysis of a century of U.S. presidential elections. PLoS One **12**(5), 1–30 (2017). https://doi.org/10.1371/journal.pone.0177970
9. Slattery, R.E., McHardy, R.R., Bairathi, R.: On the Topology of the Facebook Page Network, 3 (2013). http://arxiv.org/abs/1307.2189
10. McSweeney, P.: Gephi Network Statistics. Google Summer of Code, 1–8. (2009). http://gephi.org/google-soc/gephi-netalgo.pdf
11. Freeman, L.C.: A Set of Measures of Centrality Based on Betweenness. Sociometry. https://doi.org/10.2307/3033543
12. Bonacich, P.: Power and centrality: a family of measures. Am. J. Sociol. **92**(5), 1170–1182 (1987). https://doi.org/10.1086/228631. Accessed 1977
13. Goel, S., Anderson, A., Hofman, J., Watts, D.J.: The Structural Virality of Online Diffusion. Manag. Sci. **62**(1), 150722112809007 (2015). https://doi.org/10.1287/mnsc.2015.2158
14. Consulted in http://www.infoeleccionesmexico.com/encuestas-presidenciales-mexico.php. Accessed 21 Jan 2018

Visualizing Urban vs. Rural Sentiments in Real-Time

Jackson Howell[1(✉)] and Nathan Melenbrink[2]

[1] Harvard University Graduate School of Design, Cambridge, MA 02138, USA
jacksonhowell@gsd.harvard.edu
[2] Harvard University Wyss Institute for Biologically Inspired Engineering, Cambridge, MA 02138, USA
nathan.melenbrink@wyss.harvard.edu

Abstract. Discrepancies in sentiment between urban and rural communities represent a divide which has garnered much media attention yet so far has yielded little research or analysis. In this research, we use sentiment analysis to parse tweets in order to reveal the mood of each demographic group when discussing specific topics. We expose this method through a publicly accessible web application for sentiment tracking. Users are able to track specific keywords on Twitter in order to collect data at different scales, filtering by country, state, or even neighborhood. Using this tool, we find that across a broad range of topics generally believed to be polarizing, urban and rural groups actually express very similar sentiment scores. These results suggest that even though two demographic groups might hold completely opposite views on an issue, there is usually a certain symmetry in the emotion that both groups bring to the discourse.

Keywords: Sentiment analysis · Urban-rural divide · NLP

1 Introduction

Sentiment analysis is a form of natural language processing that can be employed to analyze the content of individual messages in large data sets, such as those gathered from social media. A common strategy involves comparing the individual words contained in a message against an index of frequently occurring keywords, where each keyword has been scored for psychological valence, corresponding to perceived negativity or positivity.

While there is a likelihood that the sentiment score of an individual text sample might be misleading (e.g. due to double negatives or unusual grammar structures) [21], sentiment scores are averaged over large numbers of text samples, from which trends can be detected and tested for significance. Mitchell et al. provide the analogy of taking the temperature in a room; the motion of a few particles cannot be expected to represent much of anything. However, an average over a sufficiently large collection of particles ultimately defines a durable quantity [19].

A. J. Morales et al. (Eds.): ICCS 2018, SPCOM, pp. 414–423, 2018.
https://doi.org/10.1007/978-3-319-96661-8_43

Previous work has used Twitter as a platform for accessing text samples that can be analyzed for sentiment. However, much of this work has dealt with static data sets [8] and focused mostly on general sentiment (mood), as opposed to querying sentiment related to a specific keyword. If there was a tool with which media outlets or social scientists could readily acquire data regarding sentiment related to specific keywords, they could better understand how an audience feels about a particular topic. If this was extended via demographic information associated with the text samples, they could study how specific communities feel about certain topics. For example, the perceived differences between urban and rural populations is a topic frequently highlighted by the media, so much so that it is often referred to as "the urban-rural divide". While there have been reports and polls that attempt to quantify and explain the widening cultural rift, there is little research concerning discrepancies in sentiment between urban and non-urban communities.

In this paper, we present a web application that addresses the need for a readily-accessible keyword-specific sentiment analysis tool. The utility of this tool is then evaluated, first through an analysis of accumulated (static) data in order to establish a baseline of sentiment differences between urban and rural populations. Finally, we use the application to track trending keywords in real-time and discuss the findings.

2 Background

Prior work in Natural Language Processing has used Twitter as a platform from which to mine data, likely due to its limited restrictions on data collection and its emphasis on short text samples (as opposed to other social media platforms that emphasize imagery or mixed-media) [5,8,24]. Recent studies using curated static data sets have combined sentiment analysis and mapping to examine geospatial and temporal trends at urban scales. França, et al. utilized tweets collected from the greater New York City area over a four month period to examine patterns of weekly activity, relating the collective geographical and temporal patterns of Twitter usage to urban activities [8]. Cao et al. analyzed tweets collected over the course of a year in the state of Massachusetts, examining sentiment relative to land use classification (e.g. commercial, farmland, industrial, residential), in an effort to investigate the holistic influence of land use and time period on public sentiment [5]. Roberts et al. utilized tweets collected over the course of a year from 60 urban green spaces in Birmingham, UK, analyzing the spatial and temporal patterns of individuals' emotions as they used these spaces typically considered valuable to urban populations in terms of mental and emotional well-being [23]. In addition to utilizing preobtained data sets collected over several months to a year, these studies are concerned with general sentiment as opposed to sentiment related to a particular topic.

Other studies have explored real-time data collection from Twitter in correlation to current events such as natural disasters [3,24], disease surveillance [16], and elections [11,28]. However, only the study by Wang et al. is concerned with sentiment analysis, providing a tool for real-time analysis of public sentiment toward presidential candidates in the 2012 U.S. election. This tool does filter the Twitter stream via keyword and could be extended to subjects other than elections; however, it does not provide any user-interface functionality for tracking different keywords and lacks a geospatial component.

Tsou et al. provide a robust interactive application, including real-time display of geolocated tweets within a target area as well as spatial, text, and temporal search functions [26]. Similarly, there are some commercial web products that allow for users to map tweets by keyword and location [1,2]. However, none of these applications provide sentiment analysis, nor do they provide users any means of understanding demographic trends such as the urban-rural divide.

Previous studies have sought to shed light on the urban-rural divide across a number of metrics, including happiness/well-being levels [4], education levels [15], obesity rates [14] and gun violence [17]. Meanwhile, the 2016 presidential election has brought renewed attention to the urban-rural divide, as voter data was more polarized along urban-rural lines than in other recent elections [9,27]. Though not strictly related to the urban-rural divide, a couple of academic studies compare urban areas with larger populations versus those with smaller populations. Hollander and Renski found that sentiment in declining cities does not differ in a statistically significant manner from stable and growing cities, suggesting that real opportunities exist to better understand urban attitudes through sentiment analysis of Twitter data [12]. Mitchell et al. look at happiness in different urban areas, comparing sentiment analysis results with other indicators of well-being [19]. They found that areas with higher numbers of tweets per capita tend to have less positive sentiment overall.

3 Methodology

In the pursuit of the objective of uncovering differences in sentiment between two demographics, we developed an application that filters live tweets by specific keyword(s) and geographic location (Fig. 1). The possible applications of such a visual analytics tool are many, from surveying political opinions, to market research, to identifying factors in the built environment that impact quality of life enough to merit an emotional response.

Only tweets containing user-supplied geolocation (which can be less than 1% of the feed) are retained. While the Twitter API makes it possible to default to the location contained in the user's profile, this information is often unreliable or incomplete. For our purposes, acceptable tweets must have location information that allows for making a determination of urban or rural status. The "place type" attribute must equal "city" as opposed to "admin", which only contains the State and Country. Tweets that contain a "city" level of location data detail can then be separated into urban and rural categories.

Tweets are then cross-referenced against a list of cities with population over 50,000 to determine if the tweet came from an urban or non-urban area, in keeping with the US Census Bureau's definition of urban areas [22]. While there

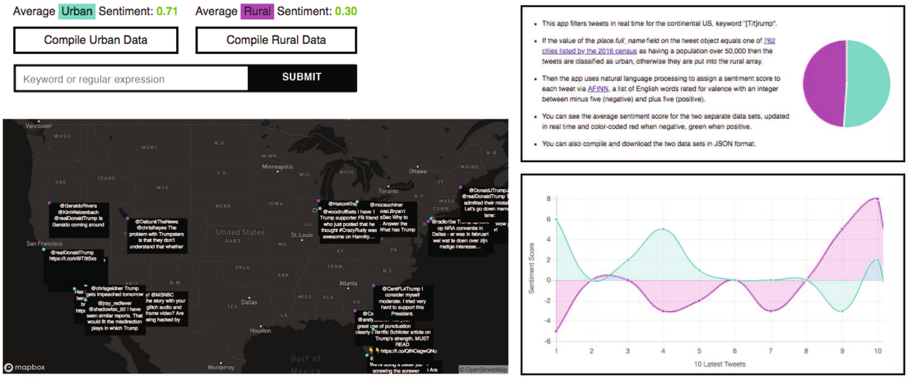


Fig. 1. The web application contains an input field for keyword(s) to be tracked. Once the user clicks submit, all tweets within the bounding box of the continental US are analyzed for sentiment and categorized as urban or rural. Visualizations show the geographic location of the tweet, as well as a running average of sentiment scores. Users are able to compile and download the collected data at any point. The application is available at https://keyword-sentiment-map.herokuapp.com

remains continuous debate over an adequate population size threshold between rural and urban places, it remains a standard to apply the threshold of 50,000 [6].

Finally, each tweet is assigned a sentiment score using the node.js Sentiment package [25], which determines a score based on the number of tokens contained in the text. A token is defined as any word or emoji. Each token is analyzed for sentiment using either Emoji Sentiment Ranking [20] or the AFINN-165 wordlist (a list of English words rated for psychological valence with an integer in the range [−5, 5]). Most words in typical written English do not appear on the wordlist and therefore receive a score of zero. The score for all tokens is summed for a cumulative sentiment score. Finally, the cumulative sentiment score is divided by the number of tokens, yielding the comparative score. While the comparative score has a theoretical range of [−5, 5], in practice most text samples receive a score in the range of [−1, 1]. This is a more balanced metric than the cumulative score as it is not biased by the length of the text, and was the primary metric used for this study. Using a library provided by De Smedt and Daelemans [7], we also recorded a polarity score and subjectivity score for each tweet, which is used for further insight.

Any study relying on Twitter should take into account that social media data cannot be considered a truly random sample of the population; some demographics may be underrepresented, especially those with limited access to smartphones and computers, as well as those whose use of social media is limited by socioeconomic, linguistic and cultural factors [12]. In 2016 the Pew Research Center found that only 24% of all online adults in the U.S (21% of all Americans) use Twitter [10]. A closer look reveals that 36% of online adults between the ages 18–29 use the platform, compared with only 10% of those 65 and over [10].

However, it is also worth noting that other disparities have mitigated with time. For example, in the past researchers have found evidence that most Twitter users were likely to be male and live in densely populated areas [18]; however, Pew's latest study reports that now 25% of adult women online use Twitter compared to 24% of adult men. Likewise, the gap narrowed between online adult users in urban and rural areas to only 26% and 24%, respectively [10].

4 Accumulated Data

In order to establish a baseline understanding of sentiment differences between urban and rural populations, we first looked at accumulated static data. A data set was compiled from tweets within the continental United States, filtering for only geolocated tweets, over the course of April 21st–26th, 2018. Ultimately a collection of one million tweets, randomly sampled from this time period, were used to establish a baseline. While a number of open-source collections of Twitter data contain more than one million tweets, they are typically not filtered for geolocated data and therefore have far fewer than one million usable data points. Furthermore, many available data sets were collected by tracking particular keywords, and therefore represent a less accurate representation of dialogue on Twitter than a random sampling method. Of all tweets collected, 40.0% were rural while 60.0% were urban. The comparative sentiment score revealed only a marginal difference between average urban (0.040) and rural (0.044) sentiments. The results of the pattern.en library found that there was no significant difference in polarity or subjectivity between the urban and rural samples [20].

We then compiled a list of 130 keywords using the Washington Post-Kaiser Family Foundation poll on urban and rural America as a guide [9]. The keywords cover dimensions of politics, economy, religion, education, sports and leisure. Then, for each keyword, all tweets containing that keyword (including sub-string matches) were extracted from the collection of one million tweets, analyzed for sentiment and categorized as either urban or rural. The number of extracted tweets depends on the frequency of the appearance of the keyword, and ranged from 251 tweets for the keyword CLIMATE to 39,636 tweets for the keyword JOB. Finally, for each keyword, the sentiment score distributions for the two data sets were compared for statistical significance.

For each keyword, we adopt the null hypothesis that both samples are drawn from the same underlying distribution function. The hypothesis is evaluated using the Kolmogorov-Smirnov test (KS test)

$$D_{n,m} = \max_x |F_{1,n}(x) - F_{2,m}(x)| \tag{1}$$

where where $F_{1,n}(x)$ and $F_{2,m}(x)$ are the empirical distribution functions for the two samples. At $\alpha = 0.005$, for each keyword the null hypothesis is rejected when

$$D_{n,m} > 1.73\sqrt{\frac{n+m}{nm}} \tag{2}$$

where m and n are the respective sample sizes [13].

Of the 130 keywords evaluated, only 10 revealed a significant difference between their sentiment distributions: PRAYER, GOD, BIBLE, BLESS, FAMILY, SAFE, NEWS, AMERICA, CITY, and PAST, all of which revealed a higher sentiment from rural samples. Figure 2 shows some of the more interesting results. The majority of keywords revealed similar distributions for both groups. The values of polarity and subjectivity, though not the primary metric of this study, were observed to loosely correlate with the comparative sentiment score; while most of the 130 keywords showed no difference in polarity or subjectivity, the 10 words that revealed a significant difference in sentiment also showed a significant difference in both polarity and subjectivity.

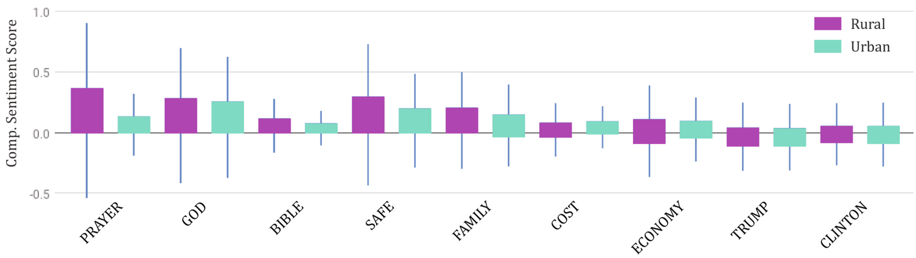


Fig. 2. Plot of the distributions of sentiment for given keywords between urban and rural populations. The y-axis shows a range in comparative sentiment scores from negative to positive, where 0 represents neutral sentiment.

Rural sentiments for messages containing the words GOD, BIBLE, BLESS or PRAYER were more positive than urban sentiments. Other keywords associated with the Christian faith (CHRIST, CHRISTIAN, JESUS) were found to be more positive among rural tweets, but not significantly so. Keywords specifically related to Islam or Judaism revealed no significant differences.

Sentiments for keywords related to political figures such as TRUMP or CLINTON were similar between both groups. More general terms such as DEMOCRAT, REPUBLICAN, LIBERAL, CONSERVATIVE, GOVERNMENT and POLITIC* all failed to reveal significant differences. The keywords ECONOMY and COST averaged more positive for the urban sample, though not significantly so. Little difference was found between rural and urban samples for other keywords related to the economy (TAX, JOBS, WORK, POOR, RICH), as well as for keywords related to immigration, crime, climate, gun control, sports, and media outlets.

5 Real-Time Data

The interface that was developed also allows for real-time data collection and analysis. This would enable a journalist, researcher or student to filter tweets and track sentiment in real-time, as it relates to unfolding events. For example, we observed from the accumulated data that keywords related to religion were among the most likely to reveal a divide between rural and urban sentiments. With this in mind, a journalist may seek to understand the variation in sentiment over the course of the National Day of Prayer, which was observed on May 3rd, 2018. While #NationalDayOfPrayer became a top trending keyword for this date, it was largely absent from the accumulated data. However, the accumulated data for the keyword PRAYER indicates that a reasonable expectation would be an average urban sentiment score of 0.05 and an average rural sentiment score of 0.18.

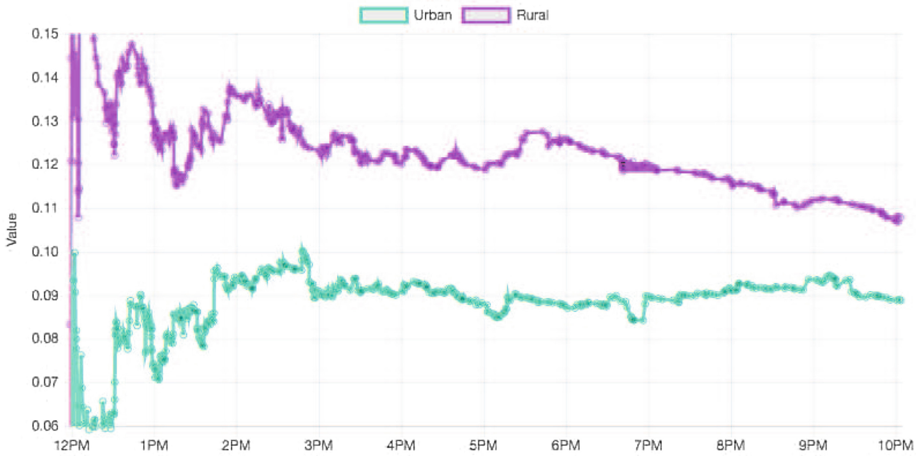


Fig. 3. Average comparative sentiment scores over time for #NationalDayofPrayer

We collected 1,062 tweets containing the keyword NATIONALDAYOFPRAYER from just before noon to shortly after 10pm, classifying 607 as urban and 455 as rural. The average urban comparative sentiment score was 0.088, compared to a rural average of 0.108. Compared to the expected results for PRAYER, we observe that there is a considerably smaller difference in sentiment for the collected sample. Closer inspection of the accumulated data for the keyword PRAYER reveals a substantial volume of positive rural tweets in the early morning hours, which was not captured by the recorded sample and may explain why the average rural sentiment score is lower than expected (Fig. 3).

On the other hand, the urban average sentiment score is higher than the accumulated average for the keyword PRAYER. While there could be any number of possible explanations for this, inspection of the data seems to reveal more instances of sarcasm and facetiousness in the urban collection. Meanwhile, genuine expressions of religious sentiment and gratitude seemed more common in the rural tweets, with users frequently retweeting statements made by local news outlets or religious organizations, or most commonly the president's first tweet on the subject at 9:27am.

6 Discussion

This research presents a visual analytics tool developed for the purpose of allowing users to better understand the urban-rural divide through sentiment analysis. We discovered through our preliminary investigation of the accumulated data that few keywords (10 out of 130) signalled a significant difference in sentiment between urban and rural populations. However, it should be understood that this does not mean that urban and rural demographics necessarily agree on the issues reflected by the keywords. For example, the fact that both groups exhibit similar sentiment scores in messages containing the word "gun" does not at all suggest that both groups share similar opinions on gun control or gun violence. However, it does suggest that the mood or emotion exhibited by each group when discussing this issue is symmetric.

Out of over a hundred sampled keywords, the only ones that were found to exhibit a significant difference between urban and rural demographics were PRAYER, GOD, BIBLE, BLESS, FAMILY, SAFE, NEWS, AMERICA, CITY, and PAST, all of which revealed a higher average sentiment from rural samples. The results suggest that people in rural areas express a more positive sentiment when they are talking about their faith and family, which is consistent with findings from polling [9]. Prior polling and reporting might foster an expectation that keywords having to do with politics, the economy, immigration, or gun laws would exhibit different sentiment scores between urban and rural demographics. However, this was not supported in our findings, suggesting that factors that are known to polarize political opinions reveal little variance in sentiment.

Often, the most impassioned text samples receive very low sentiment scores due to the inclusion of profanity and other emphatic words. It is generally understood that sentiment analysis might fail to accurately capture the sentiment for a certain percentage of tweets. However, the findings of the real-time tracking of the keyword NationalDayOfPrayer suggest that even averaging large numbers of sentiment scores can be wholly errant because the process fails to identify widespread facetiousness throughout the text samples.

While our study adopted the long-held US Census protocol for determining urbanity, in 2000, the Census Bureau expanded this classification to include two types of urban areas: Urbanized Areas (UAs) of 50,000 or more people; Urban Clusters (UCs) of at least 2,500 and less than 50,000 people [22]. We did not consider this extra category, as very few tweets came from areas with a population less than 2,500. On the other hand, the Post-Kaiser poll considered residents of counties that are part of cities with populations greater than 1 million to be urban, and also evaluated a suburban category [9]. If it seems to be the case that the Post-Kaiser methodology is more successful at identifying the cultural rifts between demographics, it would raise questions about the utility of both the former and current methods used by the US Census Bureau.

7 Future Work

While sentiment analysis can reveal the mood or emotional state that an individual exhibits while discussing a certain keyword or topic, it does not reveal that person's sentiment towards the topic. More advanced techniques such as feature identification through topic modeling or deep learning might provide better insight into the difference in sentiment regarding specific keywords.

In future work, we intend to apply the urban-suburban-rural classification method used in the Post-Kaiser survey, with the expectation that this will reveal greater variation between groups. In this paper, we searched for sentiment variance between two populations based on a wide-ranging yet fairly arbitrary set of keywords. In future work, we intend to automate a search process such that new keywords will be identified based on keywords previously found to represent a significant difference in sentiment between groups. While we focused on comparing sentiment between urban and rural populations, this approach could be more broadly adopted to compare sentiment between any two or more demographic groups.

References

1. Mapd tweetmap. https://www.mapd.com/demos/tweetmap/
2. The one million tweet map. https://onemilliontweetmap.com/
3. Avvenuti, M., Cresci, S., Marchetti, A., Meletti, C., Tesconi, M.: Ears (earthquake alert and report system): a real time decision support system for earthquake crisis management. In: Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 1749–1758. ACM (2014)
4. Berry, B.J., Okulicz-Kozaryn, A.: An urban-rural happiness gradient. Urban Geogr. **32**(6), 871–883 (2011)
5. Cao, X., MacNaughton, P., Deng, Z., Yin, J., Zhang, X., Allen, J.G.: Using twitter to better understand the spatiotemporal patterns of public sentiment: a case study in Massachusetts, USA. Int. J. Environ. Res. Public Health **15**(2), 250 (2018)
6. Cromartie, J., Bucholtz, S.: Defining the "rural" in rural america. Amber Waves **6**(3), 28 (2008)
7. De Smedt, T., Daelemans, W.: pattern.en, April 2018. https://www.clips.uantwerpen.be/pages/pattern-en
8. França, U., Sayama, H., McSwiggen, C., Daneshvar, R., Bar-Yam, Y.: Visualizing the "heartbeat" of a city with tweets. Complexity **21**(6), 280–287 (2016)
9. Gamio, L.: Urban and rural america are becoming increasingly polarized. The Washington Post (2016)
10. Greenwood, S., Perrin, A., Duggan, M.: Social media update 2016. Pew Res. Cent. **11**, 83 (2016)
11. Hamling, T., Agrawal, A.: Sentiment analysis of tweets to gain insights into the 2016 US election. Columbia Undergrad. Sci. J. **11**, 34–42 (2017)
12. Hollander, J.B., Renski, H.: Measuring Urban Attitudes Using Twitter: An Exploratory Study. Lincoln Institute of Land Policy, Cambridge (2015)
13. Isella, A.: Critical values for the two-sample kolmogorov-smirnov test (2-sided). http://sparky.rice.edu/astr360/kstest.pdf

14. Jackson, J.E., Doescher, M.P., Jerant, A.F., Hart, L.G.: A national study of obesity prevalence and trends by type of rural county. J. Rural Health **21**(2), 140–148 (2005)
15. Koricich, A., Chen, X., Hughes, R.P.: Understanding the effects of rurality and socioeconomic status on college attendance and institutional choice in the united states. Rev. High. Educ. **41**(2), 281–305 (2018)
16. Lee, K., Agrawal, A., Choudhary, A.: Real-time disease surveillance using twitter data: demonstration on flu and cancer. In: Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery And Data Mining, pp. 1474–1477. ACM (2013)
17. Lynch, K.R., Logan, T., Jackson, D.B.: "People will bury their guns before they surrender them": implementing domestic violence gun control in rural, appalachian versus urban communities. Rural Sociol. **83**, 315–346 (2018)
18. Mislove, A., Jørgensen, S., Ahn, Y.Y., Onnela, J.P., Rosenquist, J.: Understanding the demographics of twitter users, pp. 554–557. AAAI Press (2011)
19. Mitchell, L., Frank, M.R., Harris, K.D., Dodds, P.S., Danforth, C.M.: The geography of happiness: connecting twitter sentiment and expression, demographics, and objective characteristics of place. PloS One **8**(5), e64417 (2013)
20. Novak, P.K., Smailović, J., Sluban, B., Mozetič, I.: Sentiment of emojis. PloS One **10**(12), e0144296 (2015)
21. Pang, B., Lee, L., et al.: Opinion mining and sentiment analysis. Found. Trends® Inf. Retr. **2**(1–2), 1–135 (2008)
22. Ratcliffe, M., Burd, C., Holder, K., Fields, A.: Defining rural at the us census bureau. United States Census Bureau (2016)
23. Roberts, H., Sadler, J., Chapman, L.: The value of twitter data for determining the emotional responses of people to urban green spaces: a case study and critical evaluation. Urban Stud. (2018). https://doi.org/10.1177/0042098017748544
24. Sakaki, T., Okazaki, M., Matsuo, Y.: Earthquake shakes twitter users: real-time event detection by social sensors. In: Proceedings of the 19th International Conference on World Wide Web, pp. 851–860. ACM (2010)
25. Sliwinski, A.: sentiment: Afinn-based sentiment analysis for node.js, April 2018. https://github.com/thisandagain/sentiment
26. Tsou, M.H., Jung, C.T., Allen, C., Yang, J.A., Han, S.Y., Spitzberg, B.H., Dozier, J.: Building a real-time geo-targeted event observation (geo) viewer for disaster management and situation awareness. In: International Cartographic Conference, pp. 85–98. Springer (2017)
27. Ulrich-Schad, J.D., Duncan, C.M.: People and places left behind: work, culture and politics in the rural united states. J. Peasant Stud. **45**(1), 59–79 (2018)
28. Wang, H., Can, D., Kazemzadeh, A., Bar, F., Narayanan, S.: A system for real-time twitter sentiment analysis of 2012 US presidential election cycle. In: Proceedings of the ACL 2012 System Demonstrations, pp. 115–120. Association for Computational Linguistics (2012)

Urban Security Analysis in the City of Bogotá Using Complex Networks

André Ferreira[1(✉)], Guillermo Rubiano[2], and Eduardo Mojica-Nava[2]

[1] Instituto Superior Técnico, Universidade de Lisboa,
Av. Rovisco Pais 1, 1049-001 Lisbon, Portugal
andre.c.n.ferreira@tecnico.ulisboa.pt
[2] Universidad Nacional de Colombia, Cra 45, Bogotá, Colombia
{garubianog,eamojican}@unal.edu.co

Abstract. In an increasingly globalized and borderless world, fast access to reliable information about cities has become almost a necessity. From tourism to business trips and emigration, one should have a good knowledge about the destination to avoid problems and to ensure a good adaptation to the local region. As such, by exploring complex networks concepts and open data initiatives, this study focuses on the city of Bogotá as a model for security analysis, defined by official crime records and social strata percentages. In addition, a comparison of the previous data can be made with the location of police stations, as well as a urban traffic analysis. Finally, it's possible to do a regional quality classification and a quicker and safer route recommendation, in function of the reliable data extracted from databases, obtained from specialized institutions, with national accreditation for this work.

Keywords: Data science · Big data · Criminal records
Network models · Network visualization · Open data
Regional quality classification · Safety analysis · Social groups
Urban network · Urban security · Urban traffic · Complex systems

1 Introduction

The last decades have seen the reshaping of the society as information oriented. Right now, the computer is a core part of our lifes, smartphones are an extension of the human body and internet is seen as a basic need. Every document, report, event, media is digitalized, being readily accessible in a form of a PDF, a web page, a video, a database. By having the majority of the population online, constantly using and publishing more data, creation of detailed studies and impactful applications is facilitated. Beyond innovations such as online shared knowledge and social networks, this abundance of information allows the rise of data science, bringing with it the possibility to do profound analysis of the complex systems that govern the functioning of social, biological and mechanical environments. A set of complex systems that have a considerable relevance

A. J. Morales et al. (Eds.): ICCS 2018, SPCOM, pp. 424–438, 2018.
https://doi.org/10.1007/978-3-319-96661-8_44

is that of urban analysis. As cities continue to grow and absorb an increasingly large chunk of the world's population, the interest for studies regarding each city goes up. As such, nowadays it's possible to find analysis regarding the touristic value and landmarks, the average health quality, the university rankings, the fastest routes to travel through the city, the economic and social status, among others. However, it's not straightforward to find objective information concerning urban security. Although there's an unprecedented quantity of data from all around the world available online, when searching for factual safety clues about a city, there's usually only personal opinions available, that don't allow for an unbiased and detailed representation of the real system. There are academic papers that attempt to tackle this situation, as will be addressed in the related work section. Having said that, these tend to lack a clear visualization of a complete safety status of a metropolis, at a neighborhood level.

In this paper, an urban safety and traffic analysis of the city of Bogotá is presented, showing an intuitive and detailed representation of the obtained data. Bogotá is the capital of Colombia, has an estimated 8 million inhabitants and, to our best knowledge, never had a urban security report published online. As such, this represented an interesting challenge, considering the size of the city, and consequently the amount of data needed, as well as an opportunity to do an unprecedented security study, in a city that has a violent past. To address this project, we need to start by figuring out how to get a data structure that could allow us to have a geographically accurate representation of the city. Considering the generic layout of cities as streets that have intersection points, it becomes clear that a network graph, which is also used as the data format in other studies [7,18] and in services such as Google Maps, is an adequate solution. Then, we need to be able to actually create a network variable that contains the streets structure of Bogotá, as well as coordinates. After this is done, reliable and relevant data must be fetched in order to be able to do an objective safety and traffic analysis. When the graph and the data are obtained, it's a matter of carefully weighting the variables to obtain safety and traffic values for each node and edge, visualize this information in intuitive ways and derive conclusions from it.

2 Related Work

To this date, a series of studies related to urban security and traffic have been published. This is due to the increasing relevance of these two problems. With the growth of cities into big metropolis, bringing socioeconomic inequalities and a greater number of vehicles that move through the streets, there's an increase in criminality and congestion. By modeling cities as complex systems through graph theory, researchers have been able to, in some way, address these issues.

Ciccato and Uittenbogaard [3] worked on an interesting analysis on clusters of crime in Stockholm, Sweden. The study allowed a view of the regions of the Scandinavian city with more criminal activity, divided on violence and property crimes, and did a detailed time study, comparing day and night as well as winter and summer, in terms of criminal occurrences. In spite of this, there are some drawbacks to this paper. First of all, it relies on one single information from one single source, that is the crime records of Stockholm between 2006 and 2009. This means that there are no socioeconomic aspects considered, no traffic analysis and no data regarding police stations locations. Furthermore, by using the GIS framework instead of a complex networks approach, there's only a simplified visualization of the city's neighborhoods, with no possibility to view at the street level, making it difficult to extend to further research using the same data structure and preventing the functionality of calculating the best paths around the city, through use of distance and the safety knowledge gathered.

The work of Spadon et al. [18] analyzes criminal communities of San Francisco, using a complex networks representation of the city, obtained directly from Open Street Maps [14] APIs, with no corrections to the graph. Based on spatial mappings of urban crimes, they proceed with an identification of criminal communities and classification according to different types of crimes. One of the main conclusions of the paper is that different crime types share common spaces, characterizing areas that lack strategies for crime prevention. This validates our simplification of considering crime as a whole, without differentiating into different types. Although there are some important methods used in this work which are similar to ours, such as using a complex networks representation and getting spatial criminal activity data, there are some flaws which we try to address. For instances, the safety analysis focuses on the most criminal intensive region of San Francisco, instead of analyzing the whole city with unsafety scores. The security information only contemplates a criminal activity dataset, without adding data such as socioeconomic aspects, which can have an impact on a neighborhood's safety. There is no use of police stations locations in the safety analysis and they didn't take advantage of the graph structure to integrate a best path routing, in terms of both travel time and safety. A traffic analysis is completely absent.

In a 2016 paper by Solé-Ribalta et al. [1], it's presented a model based on the emergence of critical congestion phenomena, evidenced in complex networks, which allows to analytically predict the most influential points of vehicular congestion in the urban environment, based on real traffic information. Although it's a detailed approach to traffic analysis, it doesn't cover the topic of urban safety.

Geoff [7] introduced a novel complex networks tool for urban analysis, introducing the Python package OSMnx. As a new open source tool for the investigation of road networks, where one can download, analyze and build networks, a case study is discussed concerning the city of Portland, Oregon. The study consists in comparing three neighborhoods, from both metric and topological perspectives, using OSMnx. Network features such as density, connectedness, centrality, and resilience are used. In 2018, there was a publication further testing the limits of OSMnx, by doing an extensive urban morphology analysis to 27000 US cities [2]. It's clear that both papers don't address the topics of security and traffic. However, this tool is of great help for our research, considering that by obtaining a directed graph that represents the road fabric of the city of

Bogotá, it's possible to add any type of information that facilitates the understanding and visualization of georeferenced data.

In fact, our work differs from the aforementioned because it includes specific elements, highlighted as useful from previous research, in order to unify them and integrate more variables to the problems of criminality and congestion, as is the case of the relationship of security with police checkpoints. An intuitive visualization of the whole city's street network, classified by security and by traffic, which wasn't found in any previous study, is made available. In addition, we implement a route optimization where the distance traveled is shorter and, at the same time, has less traffic jams and safer with respect to other possible paths.

3 Methodology

Considering the nature of the study as being one of data science and network science, all of the implementation and analysis segments revolved around the python programming language. With a strong community, releasing packages such as numpy, for numeric operations, pandas, for data analysis, networkx, for manipulating complex networks, among many others, Python has increasingly been establishing itself as the main tool for data scientists and, with the field's growth, the applicability in machine learning and its flexibility for other domains, have made it one of the fastest growing programming languages [15]. Furthermore, considering its readability and simplicity, it allows for a more readable code, that makes it easier for anyone to review and reuse the code, as well as allowing to do a faster code development, without having memory leaks or other possible coding issues that can delay the study. Having in mind this decision of using Python as the core tool in the methodology of the study, there are 5 main steps to get the desired data analysis:

1. City graph
2. Security data
3. Traffic data
4. Police stations locations
5. Edge weight calculation

In the following pages, each of these steps are explained in further detail.

3.1 City Graph

A graph, or a network, has two main components: edges and nodes. Nodes represent possible points of interest, data points that can have practically any positive finite dimensionality. Edges represent connections between nodes, which can either be bidirectional or unidirectional, have the option of not existing at all between two given nodes and can have an attributed weight, which can be calculated as a function of certain dimensions of the end nodes. When thinking about a city, wherever it may be in the world, there are two things that are always present: streets and their inevitable intersections. Intersections can be seen as the nodes of a city, representing points of the urban landscape, where, besides geographical coordinates, dimensions can be added to represent the levels of insecurity and traffic congestion of the surroundings. The streets themselves can be seen as the edges of a city graph, which connect pairs of intersections, if there is a road between them, is either bidirectional or unidirectional, depending on whether the road is two-ways or one-way, and has a weight, which can be calculated based on a combination of the distance between the intersections, the security and traffic characteristics of the nodes' neighborhoods. All of this makes it clear that, using the streets connections, it's possible to have a network representation of a city. But now there's the problem of how to get the geographical information of a city's road map, namely of Bogotá, and how to set the intersections and their respective road connections. To handle this task, the OSMnx [7] Python package is the right tool.

OSMnx [7] is a Python package developed by Geoff Boeing, a postdoctoral Researcher at University of California, Berkeley. With it, it's possible to download spatial geometries and construct, project, visualize, and analyze street networks from OpenStreetMap's APIs [14]. OSMnx [7] improves the graphs of OpenStreetMap by only keeping the nodes which represent true street intersections and concatenates the edges that connected the intermediate false nodes into a single edge between two intersection nodes. Besides this network simplification, by being built on top of other packages such as geopandas, networkx, and matplotlib, OSMnx [7] gives access to powerful graph manipulation, graph visualization tools and path routing techniques, while additionally implementing some automatic graph analysis such as betweenness centrality, network circuity, streets per node proportions, average edge length in meters, among others.

Although OSMnx [7] allows fetching the urban network with the right political borders just by using the city's name, this feature only works for specific cities. Since it doesn't work for Bogotá, one needs to indicate a center point of the city in geographical coordinates and select a range in meters that will select the furthest away a node can be from the center. This way, it's not possible to get the right city map, but rather a graph that contains some nodes from nearby towns. By selecting a point in the center of Bogotá and a range of 20 km, the graph from Fig. 1 is constructed.

3.2 Security Data

One of the most relevant sources of data that an urban security study can have is that of criminal activity logs. By having a factual account of all the criminal occurrences that happened in a city during a given interval of years, with a geographical location of the infringement and possibly the type of crime, it enables the making of studies that can objectively classify a city's regions according to a unsafety score. This hypothesis is widely accepted, as it can be seen being applied by various studies [3,18]. However, in order for this to work flawlessly, there would need to be (i) enough data, which means having a long time interval of crime records, (ii) a time span in which the criminal panorama remains

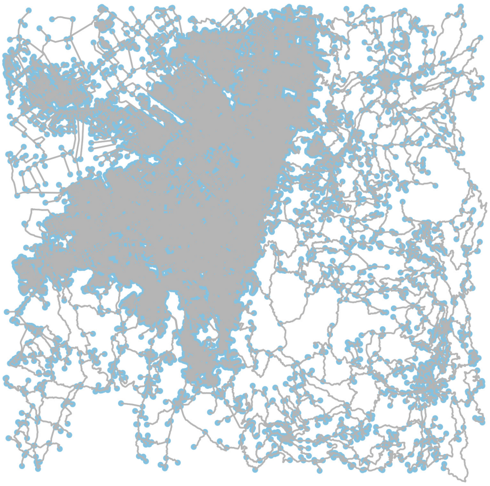


Fig. 1. Network graph representing the city of Bogotá, Colombia, as well as its surroundings, obtained with OSMnx.

sufficiently stable, as one should have data that corresponds as much as it can to the current situation, (iii) truthful data, as having crimes registered in the wrong location or even false crimes can mess up the analysis, and (iv) no missing data, as absolutely all criminal events that occurred during the dataset's time span must be registered in order for the data analysis to be perfectly representative of the real world. Considering that these requirements aren't easy to pass, that it's practically impossible to confirm a dataset's complete verification of the requirements and that this paper analysis the city of Bogotá, which has around 8 million inhabitants, making it harder to register every single crime in a dataset and validate it, we consider that there should be a better way to get the safety information.

Many studies show that there's a causal effect of poverty on crime, violence and social unrest [10,11]. As such, by having some dataset that can detail the proportion of people living in poverty conditions in each city region, it would be valuable information to determine the safety of a given region. As Bogotá has an official social stratum register, calculated through the values of the residential buildings and their corresponding terrain, it's straightforward to use a dataset from SISCRED [17], the municipality's information system, containing the number of inhabitants in each neighborhood, as well as their social stratum. In this case, for each neighborhood, we're counting on the number people from the two lowest strata and the ones without stratum, as it usually happens in Bogotá that these correspond to citizens with low income living in illegal conditions. Afterwards, we divide that number by the total amount of inhabitants of the neighborhood, resulting in the equation $Pvr_{p_i} = \frac{P_{ns_i}+P_{s_{1_i}}+P_{s_{2_i}}}{P_i}$, where Pvr_{d_i} corresponds to the poverty proportion of the population, P_{ns_i} is the number of inhabitants without stratum, $P_{s_{1_i}}$ is the number of inhabitants in stratum 1, which is the lowest in Colombian standards, $P_{s_{2_i}}$ is the number of inhabitants in stratum 2, which is the second lowest in Colombian standards, and P_i is the total amount of inhabitants, all referring to the neighborhood of node i.

Similarly, regarding the criminal activity data, we gathered a dataset from Open Data Bogotá [6], an open data initiative from the city's governance, containing criminal activity logs from 1985 until 2015. Considering that the data should be representative of the present, we filter the file to crimes committed between 2010 and 2015, which gets us 59225 criminal activities, such as assaults, bank robberies, shop robberies, violence and homicides, registered during those years. Unfortunately, the location associated to each crime is only it's locality, which is a subdivision of the city of Bogotá containing multiple neighborhoods. In spite of this, it still is useful information to determine the security level of each part of the city. As such, the way to proceed is to get the number of crimes registered in a locality, Crm_i, and divide it by the number of inhabitants of that locality, P_i, resulting in a criminal density, C_{d_i}, independent of a region's population dimension. This is equivalent to using the equation $C_{d_i} = \frac{Crm_i}{P_i}$ to determine the criminal density for each node of the graph.

Knowing that we're using a graph structure to represent the city, having the streets intersections as nodes of the graph, it's reasonable to add a new dimension to each node for a unsafety score, consisting of a weighted sum between the criminal density and the poverty proportion of the population in that node's neighborhood. As such, it's just the matter of extracting the data from each dataset's xlsx file using the pandas Python package, setting the weight parameters, k_c and k_p, and proceeding with the calculations to pin down each node's unsafety score, $Unsafety\_score_i$. Considering that the mean value of the criminal density is about 10 times lower than that of the poverty proportion, k_c should be bigger than k_p, in order to compensate for the difference of average values. This way, the influence that each unsafety factor has is more balanced.

$$Unsafety\_score_i = k_c C_{d_i} + k_p Pvr_{d_i}$$

3.3 Traffic Data

In essence, road traffic congestions are a result of having a lot of people riding cars, trucks or motorcycles, on the same pathway, at the same time. As such, the main causes for this problem can be attributed to the number of people living or commuting in a given area and to the available options for traveling between different points in the city. This last part could be addressed by evaluating characteristics such as connectivity and centrality of each node of the obtained graph, considering that a street network with a corresponding graph that has many connections should reduce traffic congestions, due to having a lot of available routes, while a high betweenness centrality value in a small set of nodes should increase traffic congestions, due to the existence of a small set of points through which a lot of people must travel through. However, we aren't going to analyze this aspect of traffic, although we'll consider it for future work. Instead, we focus on the problem related to the number of people or, more precisely, the population density. This has been shown to have a proportional relation with daily traveling delay in several studies [12,13]. Having this in consideration, a Bogotá mobility dataset from DANE [5], the national statistics institute of Colombia, is used, filtering the data to get just the population density. The dataset gives a population density value of number of inhabitants per square kilometer, for each locality. As a locality is a group of multiple neighborhoods, this doesn't allow a very detailed analysis of traffic, but does enable a broad view of possible congestion issues in certain regions of the city. Having this in consideration, the traffic score that is added to each node, $Traffic\_score_i$, is the population density of the node's locality, P_{d_i}, fetched directly from DANE's dataset [5].

$$Traffic\_score_i = P_{d_i}$$

In summary, each non police station node of the graph has the following values:

$$n_i = (Latitude, Longitude, Address, Unsafety\_score, Traffic\_score)_i$$

Where $Latitude$ and $Longitude$ are obtained from OSMnx [7], $Unsafety\_score$ and $Traffic\_score$ are calculated as shown before, and the $Address$ is the neighborhood with a central point that has the shortest distance to the node. Each neighborhood's central coordinates are found through the Google Maps geocoder, implemented in the geopy [8] Python package.

3.4 Police Stations Locations

Counting on the information of the police stations in the city, together with the crime and congestion data, one can test to see if there are possible existing correlations between the variables, verifying in what way each characteristic can be influenced by the other. For example, assuming that police stations all have comparable performance quality, we expect that, in places where the presence of police stations is higher, crime rates are lower than in locations with fewer police stations.

For information on the location of police stations in the city of Bogotá, we first experimented with Google Maps Places API, which allows to perform queries in an area defined with central geographic coordinates and a radius. As such, by performing the query for police stations through the API in Python, we should receive the names and coordinates of all the police stations. However, the amount of information returned by the medium of this service was not considerable or relevant in relation to the number of police checkpoints, as the resulting list had a maximum size of 80, which doesn't actually correspond to real life, considering the size of the city of Bogotá. For this reason, starting from the same source, Google Maps, a manual search for each of the places of interest is performed. We select these manually and store in a Google account, in the favorite places section, to have this information available when required. Finally, after downloading the coordinates and names of the police stations in a KML file, it's implemented a reading method in Python to look in the KML file for each search result and add it to our graph as police nodes. This is what makes it possible to locate the police stations and assign an attribute or specific name for the identification of the objects in the street network. As such, each node corresponding to a police station has 4 dimensions:

$$n_i = (Latitude, Longitude, Name, Type)_i$$

The *Type* dimension serves to clarify that the node is a police station.

3.5 Edge Weight Calculation

As it would be expected in a street network, OSMnx [7] utilizes short path optimization methods from networkx to calculate the route between two nodes that has the shortest distance, allowing to plot this path on the graph afterwards. The way this works is, by starting from the initial node, seeing which set of edges, that lead to the destination node, have weights such that there summation constitutes a minimum value. This means that the main parameter for calculating the best route is the edges' weights, which is just one value of one dimension of the edges in OSMnx. In essence, one can modify the edges' weights in order to change the relevant factors for optimal routing. In our point of view, there are two principal attributes that any person wants to have maximized on a route, which are fast and safe. The fast attribute is partially solved, as the default edge weight represents the distance between two intersections. However, considering the influence of traffic in transit time, the distance isn't enough to determine the fastness of a route. As such, the traffic data that was added to every node should be considered. Meanwhile, for the safe attribute, we already have a unsafety score in every node of the network. Taking all of this into account, the next logical step is to recalculate each edge's weight as a weighted sum of the distance, traffic score and unsafety score.

$$w_{ij} = k_d dist_{ij} + k_s saf_{ij} + k_t traf_{ij}$$

w represents the weight of the edge that connects nodes i and j, in which $dist_{ij}$, saf_{ij} and $traf_{ij}$ represent the distance, safety and traffic attribute of the edge, respectively, while k_d, k_s and k_t represent each dimension's weight on the calculation. The k weights are parameters that are manually tuned, according to the priority that the current user gives to the distance of the route, the safety and the traffic conditions of the neighborhoods. The remaining variables are determined using the following expressions:

$$dist_{ij} = \text{Obtained from Open Street Maps}$$

$$saf_{ij} = \frac{Unsafety\_score_i + Unsafety\_score_j}{2}$$

$$traf_{ij} = \frac{Traffic\_score_i + Traffic\_score_j}{2}$$

4 Results

In this section, the main results are presented and discussed, separating the analysis in the four topics addressed by this study.

4.1 Network and Traffic Information

As was discussed before, OSMnx [7] allows for a detailed urban morphology analysis. Using this tool, it's possible to get interesting information regarding a city viewed as a complex network. For instances, with respect to the graph obtained for the city of Bogotá, we get 89543 nodes, which represent street intersections, 236881 edges, which are every directed connection between nodes, and 133394 street segments, which are less than the edges due to considering a street in both one-way or two-way paths. Considering also the attributes regarding average street length, average streets per node and average circuity, which relates to the curvilinearity of the streets by measuring the ratio of edge lengths to the great-circle distances between the connected nodes, one can compare the complex network of Bogotá with that of some of the biggest cities in the US, using the data collected by Geoff Boeing with OSMnx [2]. As a way of doing a traffic analysis, we also added information regarding population density, from census data available on Wikipedia, and also the impact of traffic congestion, from the 2017 INRIX Global Traffic Scorecard [9], which is a study that gives a city ranking, in which a lower number corresponds to a worse congestion problem, based on an evaluation of urban travel, traffic health and vibrancy.

Table 1. Complex street networks comparison between Bogotá and some of the biggest US cities, in terms of population. Presented in descending order by population density.

City	Population density [People/km^2]	Avg. street length [m]	Avg. streets per node	Avg. circuity	2017 INRIX traffic scorecard rank
New York	11000	148	2.86	1.06	3
Boston	5368	154	2.71	1.09	14
Bogotá	5092	85	2.97	1.08	6
Miami	4866	149	2.89	1.10	10
Chicago	4594	163	2.92	1.07	17
Philadelphia	4512	159	2.87	1.08	52
Los Angeles	3275	151	2.82	1.06	1
Houston	1414	145	2.83	1.08	26

Looking at Table 1, something that is immediately noticeable is the significant difference of the average street length of Bogotá, compared to the other cities. This can be due to an alternative street pattern, as Bogotá is the only non-US city on the table, but also because the graph that was obtained for Bogotá, as was explained before, doesn't have the political borders of the city, ending up englobing some small neighboring towns like Madrid and Choachí, which can have shorter roads. Another interesting observation is that, as foreseen by other studies [12,13], there does seem to be a relation between population density and traffic congestions, as in general the most densely populated cities have a higher INRIX rankings than those less densely populated. However, there are exceptions such as Boston, Los Angeles and Houston, as they all have a ranking that is higher than a previous city on the list. Besides other factors that aren't displayed on the table, such as the fact that Bogotá doesn't have a subway transport system, Los Angeles and Houston have a lower average street length than Chicago and Philadelphia, overturning the population density differences, and both Bogotá and Miami also have a lower average street length than Boston, making them higher ranked in the traffic scorecard, despite being slightly less densely populated. There also seems to be a negative effect of low circuity on traffic, as some of the worst ranked cities of the list also have the lowest average circuity, such as the case of Los Angeles which is on a high position of the rank with New York, having a big difference in population density but sharing the lowest circuity of the cities in the table. This is probably due to the fact that a lower circuity means a more grid like street pattern, presenting more intersections, stops and traffic lights than a curvier, higher circuity street geometry.

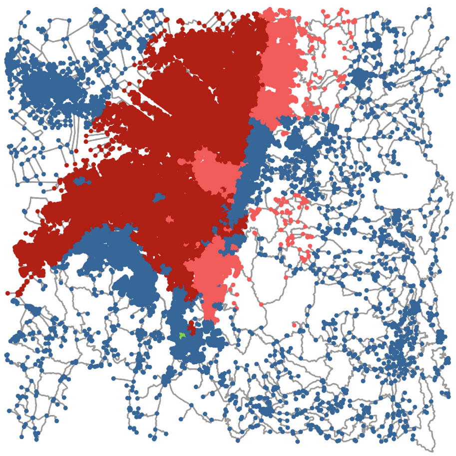


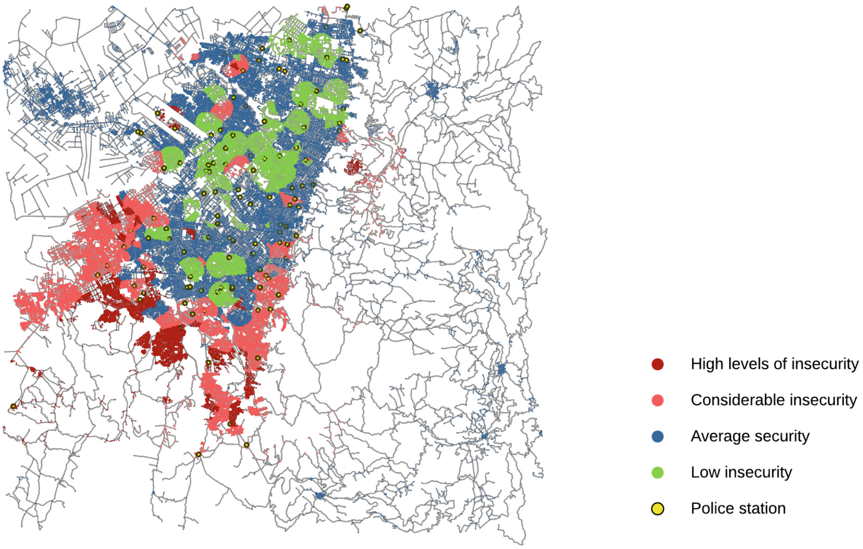
Fig. 2. Traffic data visualization on the graph of the city of Bogotá, with dark red colors representing higher congestion risk.

Figure 2 is the result of visualizing the graph nodes' traffic score, using as thresholds a population density higher than 6000 people/km^2 for the light red color and higher than 11000 for the darker red color, with all the remaining nodes having a blue color. It might seem to be a contradiction that there's such a big segment of the city with a population density higher than 11000 people/km^2 and having the census saying that the city has an average population density of only 5092 people/km^2. However, this is due to the highly changeable landscape of the Colombian capital, having modern skyscrapers but also lower densed neighborhoods, such as Chapinero, and a big rural area, with a relatively small number of inhabitants. Nevertheless, the big concentration of people in central and northern areas of Bogotá, comparable in population density to New York, make it one of the world's cities with the biggest risk and intensity of traffic jams, as mentioned in INRIX's global ranking [9].

4.2 Urban Security Graph

Using the methodology described for the security data and police stations locations, we calculate the average unsafety score of the graph's nodes and set up the thresholds of lower than $\frac{2}{9}$ the average for the green color, bigger than the average for the light red color, bigger than two times the average for the darker red color and all the rest as a blue color. The safe zone threshold is much lower than the unsafe zones thresholds due to the fact that there is a large set of nodes with a considerably high unsafety score, increasing the average score to a value that shouldn't be considered very save. The specific threshold of $\frac{2}{9}$ for safe zones is due to the fact that, due to the unsafety score being calculated in a weighted sum format, besides having less criminal logs, it should represent areas of the city with just around 7% of inhabitants in poverty conditions, which should be an indicator of a developed and safe neighborhood, in contrast with the city's average of 30%. The end result is Fig. 3, which shows a clear tendency for a safer northern region and an unsafe southern part of the city. There are also some boundary regions with a high unsafety classification, besides some more central, albeit small, regions with a moderate level of unsafety. The location of most of these zones is in accordance to the location of slums [16] and others where the economical and social conditions are at a very low standard.

With respect to the police stations locations, although one can see a strong police presence in the green regions, many of the unsafe areas displayed in red have a similar amount of police stations nearby. This means that simply implementing the same number of police stations in criminal intensive neighborhoods as applied on safer zones isn't enough. The figure suggests that, before adding up more police stations, there must be an assessment of each police station's performance, including its organization, individual quality, number of active policemen, among other aspects. Naturally, there are more variables that interfere with a region's security beyond the police services. The unsafety and poverty conditions relate to how well connected the neighborhood is to central and business points of the city, as is demonstrated in the graph by the prevalence of high unsafety scores in peripheral nodes of Bogotá's street network. As such, considering the positive example of the Colombian city of Medellín in reducing crime [4], it's safe to say that the security of Bogotá would be improved by an investment in the transport systems, mainly those that can form a stronger connection between isolated, poor regions of the city and the central, work and commercial hubs, where better social conditions, such as more jobs, education and health offers exist.



(a) (b)

Fig. 3. Network graph representing the city of Bogotá, Colombia, as well as its surroundings, having security information represented in colors.

4.3 Best Path Routing

With every node having coordinates and scores of unsafety and traffic, the edges' weights have been updated to consider these characteristics and calculate not only a fast route, but one that is safe and that tries to avoid traffic congestions. With these new weights, the optimized path between any pair of points inside the boundaries of the current graph can be determined through the networkx package's shortest path algorithm. Afterwards, OSMnx can handle the output of networkx and plot the estimated best path over the geographically accurate graph of the city.

As an example, Fig. 4 shows a route between the Virrey park and the local touristic attraction and religious site of Monserrate. The parameters used to get the edges' weights were $k_d = 1$, $k_s = 1$ and $k_t = \frac{1}{5000}$. These values were considered to allow a similar priority to all the three domains, being that the traffic values must be largely attenuated to be able to fairly compare with the other values, since it's around three orders of magnitude bigger than the distances and unsafety scores of the paths. Comparing with the security data from Fig. 3, it's possible to see the route going in a straight line through the safer green areas, while also avoiding the dangerous boundary regions of the city, and only making a sharp turn towards to the east limit, where Monserrate is located, when it's already near the destination's latitude position.

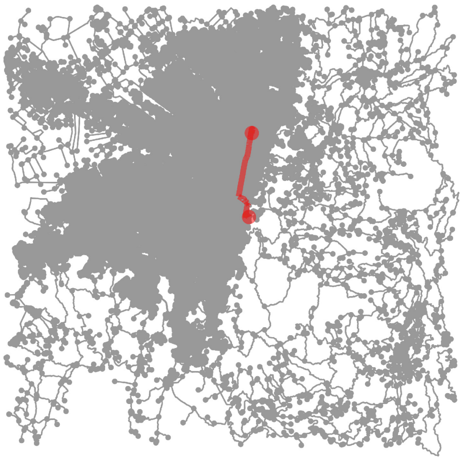


Fig. 4. Example of an optimized least distance, safer and less traffic route between Virrey park and Monserrate, in Bogotá, Colombia.

5 Conclusions

In what appears to be the first urban and traffic analysis of Bogotá, Colombia, we managed to get intuitive and visual results, that can help locals get a better understanding of their city, tourists or new immigrants get to know the safety and traffic concerns, and provide tools for political institutes to take an informed decision to fight criminal activity and improve the roadway structure. By learning from the most relevant methods applied in some of the state-of-the-art urban security and traffic studies and improving upon them with novel tools implemented on a growing programming language, we consider the main achievements of this paper to be (i) getting a highly manipulable complex network representation of Bogotá, (ii) evaluating the city's regions for traffic congestion issues through population density, with a demonstration of the validity of this method while comparing with other cities, and (iii) doing a security analysis at a neighborhood level, based on real crime events and social stratum, with clear data visualization, in a city with 8 million inhabitants. While getting these results, some interesting notes were discussed, such as a proportional effect of population density and circuity on traffic congestions and the importance of effective transport connections, between the regions struggling with poverty and the main city blocks, in the reduction of crime rates and social inequality. There was also an implementation of a optimal route search, considering safety and traffic conditions alongside the distance traveled.

Considering that some of the data had manual interference, such as the parameters in calculating each node's unsafety score, as a future work it would be interesting to improve the quality of the results by finding better parameters through machine learning algorithms. There are also the possibilities of improving the security analysis by studying the criminal behavior along the course of time and hours of the day, as well as improve the traffic analysis by considering additional variables such as local circuity, connectivity, and centrality. It would also be interesting to be able to apply similar tactics to study and compare more cities across the world.

References

1. Solé-Ribalta, A., Gómez, S., Arenas, A.: A model to identify urban traffic congestion hotspots in complex networks. R. Soc. Open Sci. **3** (2016). https://doi.org/10.1098/rsos.160098. http://www.ncbi.nlm.nih.gov/pmc/articles/PMC5098960/
2. Boeing, G.: Multi-scale analysis of 27,000 urban street networks. In: Environment and Planning B: Urban Analytics and City Science (2018)
3. Ceccato, V., Uittenbogaard, A.: space-time clusters of crime in Stockholm, Sweden. In: Review of European Studies, vol. 4. Canadian Center of Science and Education, September 2012
4. Cerdá, M., et al.: Reducing violence by transforming neighborhoods: a natural experiment in Medellín, Colombia. Am. J. Epidemiol. **175**, 1045–1053 (2012). https://doi.org/10.1093/aje/kwr428
5. Gobierno de Colombia: DANE. http://www.dane.gov.co/index.php/en/
6. Cámara de Comercio de Bogotá: Open Data Bogotá. http://datosabiertos.bogota.gov.co/
7. Geoff, B.: OSMnx: new methods for acquiring, constructing, analyzing, and visualizing complex street networks. Comput. Environ. Urban Syst. **65**, 126–139 (2017). http://geoffboeing.com/publications/osmnx-complex-street-networks/
8. Geopy Documentation. https://geopy.readthedocs.io/en/stable/
9. INRIX Global Traffic Scorecard (2017). http://inrix.com/scorecard/
10. Iyer, L., Topalova, P.: Poverty and crime: evidence from rainfall and trade shocks in India. In: Harvard Business School Working Papers (2014). https://EconPapers.repec.org/RePEc:hbs:wpaper:14-067
11. Lagi, M., Bertrand, K.Z., Bar-Yam, Y.: The food crises and political instability in North Africa and the Middle East. SSRN Electron. J. (2011)
12. Levinson, D.: Network structure and city size. PLOS ONE **7**, 1–11 (2012). https://doi.org/10.1371/journal.pone.0029721
13. Louf, R., Barthelemy, M.: How congestion shapes cities: from mobility patterns to scaling. Sci. Rep. **4** (2014). https://doi.org/10.1038/srep05561. https://www.nature.com/articles/srep05561
14. Open Street Maps. https://www.openstreetmap.org
15. Robinson, D.: Why is Python Growing So Quickly? September 2017. https://stackoverflow.blog/2017/09/14/python-growing-quickly/
16. Rueda-García, N.: The Case of Bogotá D.C., Colombia. In: Understanding Slums: Case Studies for the Global Report on Human Settlements, pp. 195–228 (2003). http://www.ucl.ac.uk/dpu-projects/Global_Report/pdfs/Bogota.pdf
17. Recreacion y Deporte Secretaria de Cultura: SISCRED. http://sispru.scrd.gov.co/siscred/
18. Spadon, G., et al.: Complex Network Tools to Understand the Behavior of Criminality in Urban Areas. University of Sao Paulo, Brazil Information Technology - New Generations, pp. 493–500, December 2016. https://link.springer.com/chapter/10.1007/978-3-319-54978-1_63

Network Analysis of ERC20 Tokens Trading on Ethereum Blockchain

Shahar Somin[1,3](✉), Goren Gordon[2,3], and Yaniv Altshuler[1,3]

[1] MIT Media Lab, Cambridge, MA, USA
{shaharso,yanival}@media.mit.edu
[2] Curiosity Lab, Industrial Engineering Department, Tel Aviv University, Tel Aviv, Israel
goren@gorengordon.com
[3] Endor Ltd., Tel Aviv, Israel

Abstract. Issuance of cryptocurrencies on top of the Blockchain system by startups and private sector companies is becoming a ubiquitous phenomenon, inducing the trading of these crypto-coins among their holders using dedicated exchanges. Apart from being a trading ledger for tokens, Blockchain can also be observed as a social network. Analyzing and modeling the dynamics of the "social signals" of this network can contribute to our understanding of this ecosystem and the forces acting within. This work is the first analysis of the network properties of the ERC20 protocol compliant crypto-coins' trading data. Considering all trading wallets as a network's nodes, and constructing its edges using buy–sell trades, we can analyze the network properties of the ERC20 network. We demonstrate that the network displays strong power-law properties, coinciding with current network theory expectations, however nonetheless, are the first scientific validation of it, for the ERC20 trading data.

The examined data is composed of over 30 million ERC20 tokens trades, performed by over 6.8 million unique wallets, lapsing over a two years period between February 2016 and February 2018.

Keywords: Complex systems · Social physics · Network analysis Blockchain · Ethereum · Smart contracts · ERC20 tokens Cryptocurrency

1 Introduction

Blockchain technology, which has been known by mostly small technological circles up until recently, is bursting throughout the globe, with a potential economic and social impact that could fundamentally alter traditional financial and social structures. Launched in July 2015 [1], the Ethereum Blockchain is a public ledger that keeps records of all Ethereum related transactions. It is shared between all participants and is based on a reward mechanism as an incentive for users to run the transactions network. A key characteristic of the Blockchain network is

A. J. Morales et al. (Eds.): ICCS 2018, SPCOM, pp. 439–450, 2018.
https://doi.org/10.1007/978-3-319-96661-8_45

its heavy reliance on cryptography to secure the transactions, addressed as the consensus mechanism. Each account consists of a public and private key duo, where the private key is used to digitally sign each account's transactions, and the public key can be used by all Blockchain participants in order to verify the transaction's validity, in a rapid, decentralized and transparent way.

The ability of the Ethereum Blockchain to store not only ownership, similarly to Bitcoin, but also execution code, in the form of *"Smart Contracts"*, has recently led to the creation of a large number of new types of "tokens", based on the Ethereum ERC20 protocol. These tokens are "minted" by a variety of players, for a variety of reasons, having all of their transactions carried out by their corresponding Smart Contracts, publicly accessible on the Ethereum Blockchain.

In this regards, the Ethereum Blockchain's transactions, and ERC20 transactions in particular, constitute a decentralized record of interactions among participants, with two interesting properties that distinguish it from most of the traditional interaction collections (such as social network activities, phone-call records, financial bank transactions):

– **Unlimited number of wallets**—The Ethereum private key mechanism enables any participant to create an unlimited amount of unique "wallets". Whereas the participant can control all of these wallets easily, it is impossible for an outside observer to explicitly associate the wallets to each other (with the exception of an implicit association, through a careful data analysis work, as can be seen in [2]). This can be compared to a mobile phone network, in which every participant may hold an infinite amount of different identities, addressed as phone numbers, all of which can be used at will. Had this property existed in reality, it would likely render most of recent seminal works in this field (such as [3–8] and many more) highly impractical, if not entirely obsolete, as demonstrated in [9].
– **Unlimited number of tokens**—The ability of participants to create not only new wallet addresses, but also an unlimited number of new *tokens* turns the Ethereum network from a single faceted means of communication of storage and execution related transactions, to a multi faceted (and in fact, an infinitely faceted) one, comprised of many different types of interactions, whose nature widely varies from payment, through decentralized trading in GPU resources [10], and to consumption of behavioral predictions [11].

As a result, the ERC20 ecosystem and the multitude of transactions it consists of, constitutes one of the most fascinating examples for decentralized networks. However, to this day there has not been any in-depth analysis of the ERC20 tokens network properties.

This work is the first attempt to analyze the ERC20 tokens through a network theory prism. We study two years of ERC20 transactions over the Ethereum Blockchain, by forming a social network from the participants and their corresponding monetary actions. We show that the ERC20 tokens data, despite being infinitely faceted and potentially comprised of unlimited amount of single-serving wallet addresses, still strongly displays several key properties known in network theory research to characterize sets of human interactions. The direct potential implication of our discovery is that the ERC20 tokens data is likely to therefore also comply with additional known network properties – leading the way for the development of an abundance of predictive and descriptive techniques for the ERC20 tokens transactions, based on known network theory oriented approaches from other domains.

The rest of the paper is organized as follows: Sect. 2 contains background on the topics of this work and review of previous literature related to it. In Sect. 3 we thoroughly describe the methodology that was used in this work, whereas the results are discussed in Sect. 4. Concluding remarks and discussion regarding future work appear in Sect. 5.

2 Background and Related Work

Blockchain's ability to process transactions in a trust-less environment, apart from trading its official cryptocurrency, the *Ether*, presents the most prominent framework for the execution of "*Smart Contracts*" [12]. Smart Contracts are computer programs, formalizing digital agreements, automatically enforced to execute any predefined conditions using the consensus mechanism of the Blockchain, without relying on a trusted authority. They empower developers to create diverse applications in a Turing Complete Programming Language, executed on the decentralized Blockchain platform, enabling the execution of any contractual agreement and enforcing its performance.

Moreover, Smart Contracts allow companies or entrepreneurs to create their own proprietary tokens on top of the Blockchain protocol [13]. These tokens are often pre-mined and sold to the public through Initial Coin Offerings (ICO) in exchange of Ether, other crypto-currencies, or *Fiat Money*. The issuance and auctioning of dedicated tokens assist the venture to crowd-fund their project's development, and in return, the ICO tokens grant contributors with a redeemable for products or services the issuer commits to supply thereafter, as well as the opportunity to gain from their possible value increase due to the project's success. The most widely used token standard is Ethereum's *ERC20* (representing Ethereum Request for Comment), issued in 2015. The protocol defines technical specifications giving developers the ability to program how new tokens will function within the Ethereum ecosystem.

This brand new market of ERC20 compliant tokens is fundamental to analyze, as it is becoming increasingly relevant to the financial world. Issuing tokens on top of the Blockchain system by startups and other private sector companies is becoming a ubiquitous phenomenon, inducing the trade of these crypto-coins to an exponential degree. Since 2017, Blockchain startups have raised over 7 Billion dollars through ICOs. Among the largest offerings, *Tezos* raised $232M for developing a smart contracts and decentralized governance platform; *Filecoin* raised $205M to deploy a decentralized file storage network; *EOS* raised over $185M to fund scalable smart contracts platform and *Bancor*, who managed to raise $153M for deploying a Blockchain-based prediction market.

Apart from being formed by countless stake-holders and numerous tokens, the ERC20 transactional data also presents full data of prices, volumes and holders distribution. This, alongside with daily transactions of anonymised individuals is otherwise scarce and hard to obtain due to confidentiality and privacy control, hence providing a rare opportunity to analyze and model financial behavior in an evolving market over a long period of time.

In the past two decades, network science has exceedingly contributed to multiple and diverse scientific disciplines. Applying network analysis and graph theory have assisted in revealing the structure and dynamics of complex systems by representing them as networks, including social networks [14–16], computer communication networks [17], biological systems [18], transportation [19,20], IOT [21], emergency detection [22] and financial trading systems [23–25].

Most of the research conducted in the Blockchain world, was concentrated in Bitcoin, spreading from theoretical foundations [26], security and fraud [27,28] to some comprehensive research in network analysis [29–31]. The world of Smart contracts has recently inspired research in aspects of design patterns, applications and security [32–35], policy towards ICOs has also been studied [13]. However, the comprehensive analysis of ERC20 tokens, with emphasis on the investigation of the transaction graph built from their related activity on the Blockchain, is still lacking. In this paper we aim to examine how this prominent field can enhance the understanding of the underlying structure of the ERC20 tokens trading data.

3 Methodology

3.1 Data

In order to preserve anonymity in the Ethereum Blockchain, personal information is omitted from all transactions. A User, represented by their wallet, can participate in the economy system through an address, which is attained by applying *Keccak-256* hash function on his public key. The Ethereum Blockchain enables users to send transactions in order to either send Ether to other wallets, create new Smart Contracts or invoke any of their functions. Since Smart Contracts are scripts residing on the Blockchain as well, they are also assigned a unique address. A Smart Contract is called by sending a transaction to its address, which triggers its independent and automatic execution, in a prescribed manner on every node in the network, according to the *data* that was included in the triggering transaction.

Smart Contracts representing ERC20 tokens comply with a protocol defining the manner in which the token is transferred between wallets and the form in which data within the token is accessed. Among these requirements, is the demand to implement a *transfer* method, which will be used for transferring the relevant token from one wallet to another. Therefore, each transfer of an ERC20 token will be manifested by a wallet sending a transaction to the relevant Smart Contract. The transaction will encompass a call to the *transfer* method in its *data* section, containing the amount being transferred and its recipient wallet.

Each such token transfer results in altering the 'token's balance', which is kept and updated in its corresponding Smart Contract's storage.

We obtain the ERC20 transactions basing on the further requirement of the ERC20 protocol, demanding that each call to the *transfer* method will be followed by sending a *Transfer* event and updating the event's logs with all relevant information regarding the token transfer. We therefore call an Ethereum full node's JSON API and fetch all logs matching to the *Transfer* event structure [36]. Parsing these logs result in the following fields per transaction: *Contract Address* - standing for the address of the Smart Contract defining the transferred token, *Value* - specifying the amount of the token being transferred, *Sender* and *Receiver* addresses, being the wallet addresses of the token's seller and buyer, correspondingly.

We have retrieved all ERC20 tokens transactions spreading between February 2016 and February 2018, resulting in $30,347,248$ transactions and $18,517$ token address. Due to the restriction on changing and tempering Smart Contracts, any modification made to a token's designated Smart Contract involves a definite change in it's associated Contract Address. As a result, a token can change addresses throughout it's lifespan, though for any point in time, it will only be assigned to a single relevant *Contract Address*. Therefore, the above mentioned amount of unique contract addresses serves merely as an upper bound to the amount of unique tokens. Since we do not restrict ourselves to a specific type of token, but observe the network as a whole trading system, this non-unique identification of tokens doesn't affect our analysis of the network.

The dataset of ERC20 tokens transactions is extremely diverse and wide-ranging, where not only any ERC20 token might correspond to multiple contract addresses, and therefore being considered as various different tokens by our analysis, but also the characteristics of the different tokens are extremely varied. For instance, the tokens differ in their age, their economic value, activity volume and number of token holders, some merely serve as test-runs, others aren't tradable in exchanges yet, and some, according to popular literature, are frauds, all residing next to actual real-world valuable tokens.

3.2 Graph Analysis

In order to perceive the network's structure and assess the connectivity of its nodes, one should examine the network's degree distribution, considering both in-degree and out-degree, indicating the number of incoming and outgoing connections, correspondingly. The degree distribution $P(k)$ signifies the probability that a randomly selected node has precisely the degree k.

In random networks of the type studied by Erdös and Rényi [37], where each edge is present or absent with equal probability, the nodes' degrees follow a *Poisson* distribution. The degree obtained by most nodes is approximately the average degree $\bar{k}$ of the network. These properties are also manifested in dynamic networks [38]. In contrast to random networks, the nodes' degrees of social networks (such as the Internet or citation networks) often follow a *power law* distribution [39]:

$$P(k) = k^{-\alpha} \tag{1}$$

The power law degree distribution indicates that there is a non-negligible number of extremely connected nodes even though the majority of nodes have small number of connections. Therefore the degree distribution has a long right tail of values that are far above the average degree. Power law distributions can be found in many real networks, Newman [16] summarized several of them, including word frequency, citations, telephone calls, web hits, or the wealth of the richest people.

4 Results

As discussed, we study an extremely diverse and wide-ranging dataset. In order to present a first glimpse on the diversity of ERC20 tokens transactional data, we explore the distribution of token popularity, in terms of buyers and sellers amount.

Definition 1. *Let CT be an ERC20 token. The token's Buying Popularity during timespan t, denoted by BP_t, is defined as the number of unique wallets which bought the token during the examined time:*

$$BP_t(CT) := |\{w_v \| \text{ wallet } w_v \text{ bought CT during time } t\}| \tag{2}$$

Correspondingly, Selling Popularity during time t, denoted by SP_t, is defined as the amount of unique wallets who sold token CT during this time:

$$SP_t(CT) := |\{w_v \| \text{ wallet } w_v \text{ sold CT during time } t\}| \tag{3}$$

As Fig. 1 reflects, ERC20 tokens' both Buying and Selling Popularities follow a power-law distribution, thereby expressing the diversity of token holders along a 2 years period, between February 2016 and February 2018. Particularly, it can be seen that most tokens are traded by an extremely small amount of users and on the other hand, a few popular tokens exist, traded by a very large group of users during the examined timespan.

We further aim to examine whether the ERC20 network satisfies the known characteristics of other real-world networks, first and foremost examining its degree distribution. We therefore construct the following directed graph, $G_{FT}(V, E)$, standing for *ERC20 Full Transactions Graph*, including all transactions in the timespan between February 2016 and February 2018. The resulting graph consists of $6,890,237$ vertices and $17,392,610$ edges.

The set of vertices V consists of all ERC20 trading wallets in this period, where any vertex u represents a trading wallet w_u. Out-going edges depict transactions in which wallet w_u sold any type of ERC20 token to other wallets, and in-coming edges to u are formed as result of transactions in which w_u bought any ERC20 token from others. Formally, $E \subseteq V \times V$ s.t.:

$$E := \{(u, v) \| \text{ wallet } w_u \text{ sold any ERC20 token to } w_v\} \tag{4}$$

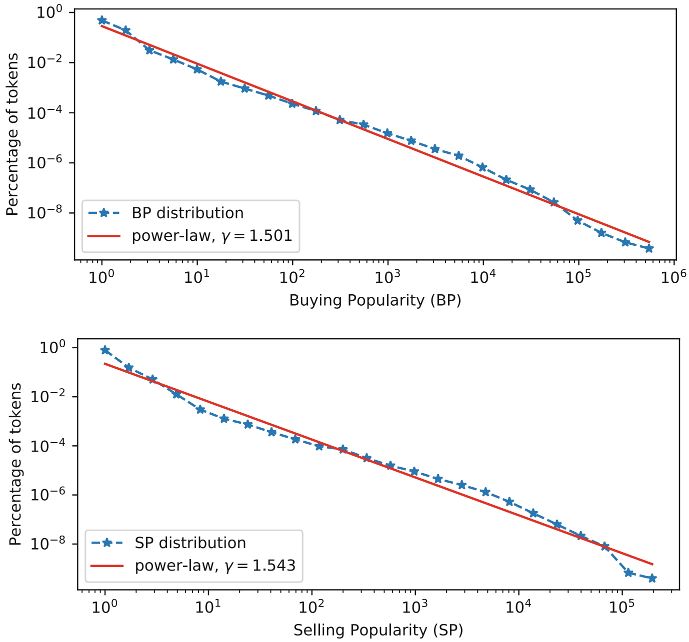


Fig. 1. Histogram of the Buying Popularity BP_t (upper panel) and Selling Popularity SP_t (lower panel) of all ERC20 Crypto-coins [see Definition 4], for t being a two years period lapsing from February 2016 to February 2018. Depicting the probability a coin would have certain popularity among buyers and sellers correspondingly, both demonstrating a power-law distribution.

Out-degree of vertex u represents the number of unique wallets buying tokens from w_u and its in-degree depicts the number of unique wallets selling tokens to it.

Surprisingly, despite the great variance between the traded tokens in the network, we discovered that the degree distribution depicts a strong power-law pattern, as presented in Fig. 2. Hence the *ERC20 Full Transactions Graph*, G_{FT}, displays similar connectedness structure to other real-world networks, such as [14–16], presenting a non-negligible number of highly connected nodes even though the majority of nodes have small number of connections, both in buying and selling transactions. We have additionally analyzed ERC20 transaction graphs based on varying length periods between 3 days to 3 months, and validated our findings across 20 different points in time. We have observed that in all cases the power-law degree distribution is preserved and presents roughly similar γ values.[1]

Elaborating on this last observation, which is uncovered for the first time in this work, it can be seen that the economic activity on the ERC20 network—both

[1] We omit these results from the current version, due to space limitations, and they will appear in a future, extended version.

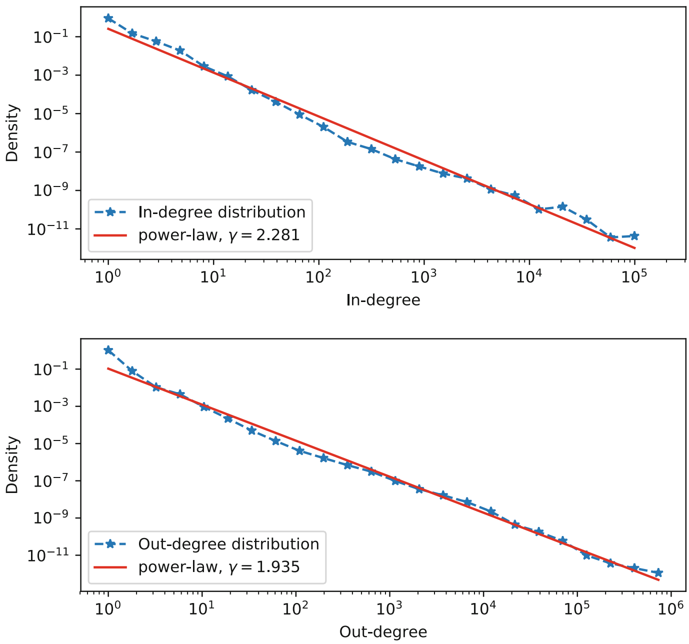


Fig. 2. Analysis of Blockchain network dynamics for a 2 years period from February 2016 to February 2018. The network's nodes represent ERC20 wallets and edges are formed by ERC20 buy-sell transactions. Outgoing degree of a node reflects the number of unique wallets receiving funds from that node, regardless of the token being transferred, and vice-versa for incoming degree. Both outgoing and incoming degrees present a power-law distribution, similarly to what was demonstrated in analysis of mobile phone, citation data and many other real-world networks [16]

outgoing, incoming, and reciprocal—converges to a heavy-tail distribution. This discovery has several important (and partially counterintuitive) consequences:

Decentralization. The first derivative from the power law phenomena demonstrated in this work, is the strengthening of the *ERC20* environment's decentralization property. Decentralization in this context is manifested by the existence of a large number of medium sized hubs, taking part in the network's activity, constituting a network that is not governed by a single major player, both in the sense of trading wallets as well as in traded tokens. Decentralization, forming a key feature of the Blockchain technology, and for some – its main "claim to fame", is both celebrated and questioned. By clearly showing the emergence of a heavy tail distribution within the trading behavior of its users, we can for the first time, provide a concrete data-driven proof for the inherent decentralization of *ERC20* tokens, which remains stable across various time-periods, and length of analysis windows.

Robustness. An immediate implication of the decentralization of the *ERC20* tokens network, is also its robustness. Several works have used percolation theory [40] to demonstrate that such network structures are often less subject to manipulations using small correlated groups [41], making it easier for the majority of the crowd to maintain relative freedom.

Diversity. Subsequently, it also facilitates the creation of new emerging tokens, as it increases the probability that they would be adopted by a non-negligible-in-size group of "first-adopters". This is specifically important for an environment that aims to provide opportunities for the fast creation and adoption of new applications.

Maturity. Several critics have referred to *Ethereum*, as well as to the *ERC20* tokens in general, as an immature economic structure, that is unstable and certainly not well representing a "normal" human economy. The stable and multi-faceted power-law patterns demonstrated in our analysis imply that these criticisms are, to the very list, partially unjustified. The convergence of tokens distributions, as well as buying and selling activities is a typical characteristic of "natural human behavior" [4] and specifically mature economies [42,43]. Furthermore, as demonstrated in works such as [44] it is also an efficient substance for the natural evolution of sub-communities.

Opportunities for future research. A strong embedding of power-law characteristics implies the likely usability of the multitude of known techniques utilizing this feature for various purposes such as anomalies detection [22], marketing optimization [45,46] and cybersecurity [47].

5 Concluding Remarks and Future Work

In this paper, we have demonstrated for the first time that the ERC20 tokens transactional data displays several properties known to be associated with networks that are comprised of human interactions, and social networks specifically. This occurs despite the fact that the Blockchain protocol enables the creation of an unlimited number of "tokens", causing diverse sub-domains to reside together over the same protocol, and regardless of an unlimited amount of wallets, resulting in different identities controlled by a single individual.

Specifically, we have modeled the transactions as a network that is comprised of wallets, connected through transactions, and found that the degree distribution of nodes in the network presents a power-law pattern. In addition, we have shown that tokens popularity among buyers and sellers also follows a power-law model. These preliminary results indicate that (somewhat surprisingly) despite its diversity, ERC20 data presents a social behavior. This leads us to explore whether other aspects of network theory can emerge from this data. Such fields include short path lengths and clustering coefficient analysis [48], centrality measures [49], connected components behavior [5] and community structure study [50]. We have already been able to demonstrate some of these phenomena using the ERC20 data as well, however they were not included in this work due to space considerations.

References

1. Buterin, V., et al.: A next-generation smart contract and decentralized application platform. White paper (2014)
2. Altshuler, Y., Elovici, Y., Cremers, A.B., Aharony, N., Pentland, A.: Security and Privacy in Social Networks. Springer Science & Business Media, New York (2012)
3. Gonzalez, M.C., Hidalgo, C.A., Barabasi, A.-L.: Understanding individual human mobility patterns. Nature **453**, 779–782 (2008)
4. Barabasi, A.: The origin of bursts and heavy tails in human dynamics. Nature **435**(7039), 207–211 (2005)
5. Candia, J., González, M.C., Wang, P., Schoenharl, T., Madey, G., Barabási, A.-L.: Uncovering individual and collective human dynamics from mobile phone records. J. Phys. A: Math. Theor. **41**(22), 224015 (2008)
6. Eagle, N., Pentland, A., Lazer, D.: Inferring social network structure using mobile phone data. Proc. Nat. Acad. Sci. (PNAS) **106**, 15274–15278 (2009)
7. Altshuler, Y., Aharony, N., Pentland, A., Elovici, Y., Cebrian, M.: Stealing reality: when criminals become data scientists (or vice versa). In: Intelligent Systems, vol. 26, pp. 22–30. IEEE, November–December 2011
8. Onnela, J.-P., Saramäki, J., Hyvönen, J., Szabó, G., Lazer, D., Kaski, K., Kertész, J., Barabási, A.-L.: Structure and tie strengths in mobile communication networks. Proc. Nat. Acad. Sci. **104**(18), 7332–7336 (2007)
9. Altshuler, Y., Aharony, N., Fire, M., Elovici, Y., Pentland, A.: Incremental learning with accuracy prediction of social and individual properties from mobile-phone data. CoRR (2011)
10. Golem (2017)
11. Endor – inventing the *"Google for Predictive Analytics"* (2017)
12. Wood, G.: Ethereum: a secure decentralised generalised transaction ledger. Ethereum Proj. Yellow Pap. **151**, 1–32 (2014)
13. Catalini, C., Gans, J.S.: Initial coin offerings and the value of crypto tokens, Technical report, National Bureau of Economic Research (2018)
14. Barrat, A., Barthelemy, M., Vespignani, A.: Dynamical Processes on Complex Networks. Cambridge University Press, Cambridge (2008)
15. Newman, M.E.: The structure and function of complex networks. SIAM Rev. **45**(2), 167–256 (2003)
16. Newman, M.E.: Power laws, pareto distributions and Zipf's law. Contemp. Phys. **46**(5), 323–351 (2005)
17. Pastor-Satorras, R., Vespignani, A.: Evolution and Structure of the Internet: A Statistical Physics Approach. Cambridge University Press, Cambridge (2007)
18. Barabasi, A.-L., Oltvai, Z.N.: Network biology: understanding the cell's functional organization. Nat. Rev. Genet. **5**(2), 101 (2004)
19. Shmueli, E., Mazeh, I., Radaelli, L., Pentland, A.S., Altshuler, Y.: Ride sharing: a network perspective. In: International Conference on Social Computing, Behavioral-Cultural Modeling, and Prediction, pp. 434–439. Springer (2015)
20. Altshuler, Y., Puzis, R., Elovici, Y., Bekhor, S., Pentland, A.S.: On the rationality and optimality of transportation networks defense: a network centrality approach. Secur. Transp. Syst. 35–63 (2015)
21. Altshuler, Y., Fire, M., Aharony, N., Elovici, Y., Pentland, A.: How many makes a crowd? On the correlation between groups' size and the accuracy of modeling. In: International Conference on Social Computing, Behavioral-Cultural Modeling and Prediction, pp. 43–52. Springer (2012)

22. Altshuler, Y., Fire, M., Shmueli, E., Elovici, Y., Bruckstein, A., Pentland, A.S., Lazer, D.: The social amplifier-reaction of human communities to emergencies. J. Stat. Phys. **152**(3), 399–418 (2013)
23. Altshuler, Y., Pan, W., Pentland, A.: Trends prediction using social diffusion models. In: International Conference on Social Computing, Behavioral-Cultural Modeling and Prediction, pp. 97–104. Springer (2012)
24. Pan, W., Altshuler, Y., Pentland, A.: Decoding social influence and the wisdom of the crowd in financial trading network. In: 2012 International Confernece on Privacy, Security, Risk and Trust (PASSAT) and 2012 International Confernece on Social Computing (SocialCom), pp. 203–209. IEEE (2012)
25. Shmueli, E., Altshuler, Y., et al.: Temporal dynamics of scale-free networks. In: International Conference on Social Computing, Behavioral-Cultural Modeling, and Prediction, pp. 359–366. Springer (2014)
26. Bonneau, J., Miller, A., Clark, J., Narayanan, A., Kroll, J.A., Felten, E.W.: Sok: research perspectives and challenges for bitcoin and cryptocurrencies. In: 2015 IEEE Symposium on Security and Privacy (SP), pp. 104–121. IEEE (2015)
27. Meiklejohn, S., Pomarole, M., Jordan, G., Levchenko, K., McCoy, D., Voelker, G.M., Savage, S.: A fistful of bitcoins: characterizing payments among men with no names. In: Proceedings of the 2013 Conference on Internet Measurement Conference, pp. 127–140. ACM (2013)
28. Shrobe, H., Shrier, D.L., Pentland, A.: New Solutions for Cybersecurity. MIT Press, Cambridge (2018)
29. Ron, D., Shamir, A.: Quantitative analysis of the full bitcoin transaction graph. In: International Conference on Financial Cryptography and Data Security, pp. 6–24. Springer (2013)
30. Maesa, D.D.F., Marino, A., Ricci, L.: Uncovering the bitcoin blockchain: an analysis of the full users graph. In: 2016 IEEE International Conference on Data Science and Advanced Analytics (DSAA), pp. 537–546. IEEE (2016)
31. Lischke, M., Fabian, B.: Analyzing the bitcoin network: the first four years. Future Internet **8**(1), 7 (2016)
32. Bartoletti, M., Pompianu, L.: An empirical analysis of smart contracts: platforms, applications, and design patterns. In: International Conference on Financial Cryptography and Data Security, pp. 494–509. Springer (2017)
33. Anderson, L., Holz, R., Ponomarev, A., Rimba, P., Weber, I.: New kids on the block: an analysis of modern blockchains. arXiv preprint arXiv:1606.06530 (2016)
34. Christidis, K., Devetsikiotis, M.: Blockchains and smart contracts for the internet of things. IEEE Access **4**, 2292–2303 (2016)
35. Atzei, N., Bartoletti, M., Cimoli, T.: A survey of attacks on ethereum smart contracts (SoK). In: International Conference on Principles of Security and Trust, pp. 164–186. Springer (2017)
36. Json prc api (2018)
37. Erdös, P., Rényi, A.: On random graphs, i. Publicationes Mathematicae (Debrecen) **6**, 290–297 (1959)
38. Erdos, P., Renyi, A.: On the evolution of random graphs. Publ. Math. Inst. Hung. Acad. Sci. **5**, 17–61 (1960)
39. Albert, R., Barabási, A.-L.: Statistical mechanics of complex networks. Rev, Modern Phys. **74**(1), 47 (2002)
40. Callaway, D.S., Newman, M.E., Strogatz, S.H., Watts, D.J.: Network robustness and fragility: percolation on random graphs. Phys. Rev. Lett. **85**(25), 5468 (2000)
41. Liu, Y.-Y., Slotine, J.-J., Barabási, A.-L.: Controllability of complex networks. Nature **473**(7346), 167 (2011)

42. Barabási, A.-L.: Linked: The new science of networks (2003)
43. Barabási, A.-L.: The elegant law that governs us all (2017)
44. Palla, G., Barabasi, A., Vicsek, T.: Quantifying social group evolution. Nature **446**(7136), 664–667 (2007)
45. Altshuler, Y., Shmueli, E., Zyskind, G., Lederman, O., Oliver, N., Pentland, A.: Campaign optimization through behavioral modeling and mobile network analysis. IEEE Trans. Computat. Soc. Syst. **1**(2), 121–134 (2014)
46. Altshuler, Y., Shmueli, E., Zyskind, G., Lederman, O., Oliver, N., Pentland, A.: Campaign optimization through mobility network analysis. In: Geo-Intelligence and Visualization Through Big Data Trends, pp. 33–74 (2015)
47. Pentland, A., Altshuler, Y.: Social Physics and Cybercrime. In: New Solutions for Cybersecurity, pp. 351–364. MIT Press (2018)
48. Watts, D.J., Strogatz, S.H.: Collective dynamics of 'small-world' networks. Nature **393**(6684), 440 (1998)
49. Borgatti, S.P.: Centrality and network flow. Soc. Netw. **27**(1), 55–71 (2005)
50. Girvan, M., Newman, M.E.: Community structure in social and biological networks. Proc. Natl. Acad. Sci. **99**(12), 7821–7826 (2002)

The Principle of Relatedness

César A. Hidalgo[1](✉), Pierre-Alexandre Balland[2], Ron Boschma[2,3], Mercedes Delgado[4], Maryann Feldman[5], Koen Frenken[6], Edward Glaeser[7,8], Canfei He[9], Dieter F. Kogler[10], Andrea Morrison[2], Frank Neffke[11], David Rigby[12], Scott Stern[4,8], Siqi Zheng[13,14], and Shengjun Zhu[9]

[1] Collective Learning Group, the MIT Media Lab, Massachusetts Institute of Technology, Cambridge, USA
hidalgo@mit.edu
[2] Department of Human Geography and Planning, Utrecht University, Utrecht, Netherlands
[3] Business School, University of Stavanger, Stavanger, Norway
[4] MIT Sloan School of Management, Cambridge, USA
[5] Department of Public Policy, University of North Carolina at Chapel Hill, Chapel Hill, USA
[6] Copernicus Institute of Sustainable Development, Utrecht University, Utrecht, Netherlands
[7] Department of Economics, Harvard University, Cambridge, USA
[8] National Bureau of Economic Research, Cambridge, USA
[9] College of Urban and Environmental Sciences, Peking University, Beijing, China
[10] Spatial Dynamics Lab and School of Architecture Planning and Environmental Policy, University College Dublin, Dublin, Ireland
[11] Center for International Development, Harvard University, Cambridge, USA
[12] Department of Geography and Department of Statistics, University of California, Los Angeles, USA
[13] Department of Urban Studies and Planning, MIT, Cambridge, USA
[14] School of Civil Engineering, Tsinghua University, Beijing, China

Abstract. The idea that skills, technology, and knowledge, are spatially concentrated, has a long academic tradition. Yet, only recently this hypothesis has been empirically formalized and corroborated at multiple spatial scales, for different economic activities, and for a diversity of institutional regimes. The new synthesis is an empirical principle describing the probability that a region enters—or exits—an economic activity as a function of the number of related activities present in that location. In this paper we summarize some of the recent empirical evidence that has generalized the principle of relatedness to a fact describing the entry and exit of products, industries, occupations, and technologies, at the national, regional, and metropolitan scales. We conclude by describing some of the policy implications and future avenues of research implied by this robust empirical principle.

C. A. Hidalgo and P.-A. Balland—Contributed equally.

A. J. Morales et al. (Eds.): ICCS 2018, SPCOM, pp. 451–457, 2018.
https://doi.org/10.1007/978-3-319-96661-8_46

Keywords: Economic complexity · Relatedness · Economic geography

1 Introduction

Why are human creativity and productivity concentrated in space and time? Is the computer revolution in Silicon Valley similar to England's Industrial Revolution or to Florence's Renaissance?

An old idea in economic geography is that in dense industrial clusters the secrets of a trade are, "as it were in the air." [1] James Watt's steam engine required the metal works of John "Iron Mad" Wilkinson, just like Apple's Macintosh required the idea of the Graphical User Interface (GUI) developed at Xerox's PARC. The idea that skills, technology, and knowledge, are spatially concentrated, has a long academic tradition, dating back to Johan von Thünen and Alfred Marshall in the nineteenth century, and to Harold Hotelling, Walter Christaller, August Lösch, Waldo Tobler, Jane Jacobs, and Michael Porter in the twentieth. Yet, only recently this hypothesis has been empirically formalized and corroborated at multiple spatial scales, for different economic activities, and for a variety of institutional regimes. The new synthesis is an empirical principle describing the probability that a region enters (or exits) an economic activity as a function of the number of related activities present in that location. This synthesis is *the principle of relatedness*.

2 The Principle of Relatedness

In principle, we say that two activities, such as products, industries, or research areas are related when they require similar knowledge or inputs (Fig. 1A). For instance, shirts and blouses are related because they are manufactured using similar materials and technologies. In practice, however, which inputs and knowledge are used in a production process are at best imperfectly observed. The modern methods used to *infer* relatedness are agnostic about the specific inputs or complementarities that drive relatedness among activities. They look, for instance, at the co-export of products [2], the flow of labor among industries [3], or combined measures of input-output links and shared labor pools [4]. This methodological flexibility, has allowed scholars from a variety of fields to document a robust and reproducible relationship between the probability that a location will develop expertise in a new industry [5–7], technology [8, 9], research area [10], product [2], or occupation [11], and the number of related activities that are already present in that location.

Hidalgo et al. [2] introduced the idea of the product space—a network connecting products that are likely to be exported in tandem—to show that the probability that a country will start exporting a product increases with the number of related products that this country already exports (Fig. 1B). Neffke et al. [5], and Zhu et al. [6], looked respectively, at Sweden and China to show that the probability that an industry will enter a region increases with the number of related industries present in it. Kogler et al. [8] and Boschma et al. [9] connected technology classes that co-occur in patents to

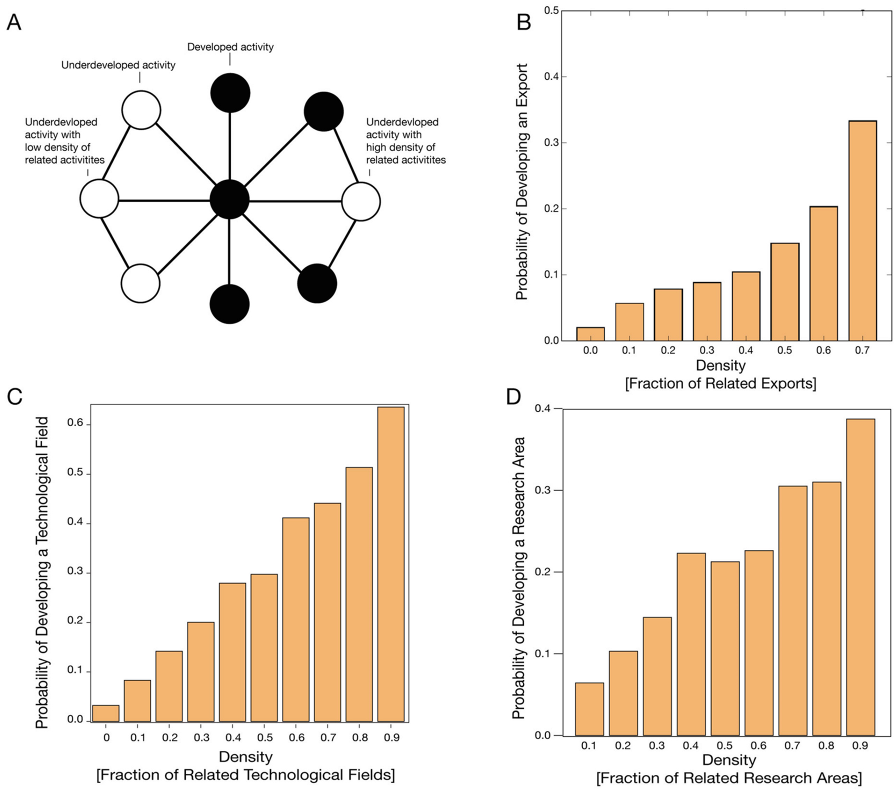


Fig. 1. **A**. The number of related activities present in a location can be measured using density: a weighted average of the fraction of related activities that are present in that location. Relatedness is represented using a network, where nodes represent activities, such as products, research areas, or technology classes, and links connect related activities (products that are co-exported, or research areas that share authors). Black nodes indicate the activities that are present in a location. The activities that are not present in that location are shown in white. The node on the far left represents an activity with low density (few related activities present in that location, 1 out of 3). The node on the far right is an activity with high density (with many related activities present in that location, 3 out of 3). **B.** Probability that a country will develop comparative advantage in a product in a four-year period (that it will export more of that product than what should be expected from that country's total exports and the size of that product's market) as a function of the density of related products (fraction of related products already exported by that country). **C.** Probability that a city (U.S. Metropolitan Statistical Area) will produce more patents in a technology class than what is expected from that city's total patenting volume and the size of the technology class as a function of the density of related technological fields. **D**. Probability that a country will develop a research area in a period of three years (that it will publish more papers than what is expected from that country's total publication volume and the size of the field) as a function of the density of related research areas.

show that cities in the United States were more likely to begin patenting in a technology class when they had expertise in related technologies (Fig. 1C). Guevara et al. [10], connected research areas to show that the probability that a scholar, university, or country, starts publishing in a new research area increases with the number of related areas in which a scholar, university, or country, has expertise (Fig. 1D).

These efforts have generalized the principle of relatedness from a simple observation, to a formal empirical result that is valid for multiple spatial scales, economic activities, and institutional backgrounds. Yet, these results have also demonstrated the empirical strength of the principle. In Figs. 1b–d, the probability of entering an activity rises between eight-fold and twenty-fold when we move from an unrelated activity to a related activity. This shows that the principle of relatedness is not only robust and ubiquitous, but also, strong.

But does the principle of relatedness teach us something about economic development? And is it a principle that matters?

The high degree of reproducibility of this principle hints at something fundamental: the variety of mechanisms by which economies and organizations learn. The principle of relatedness is not the result of trivial input output relationships, like needing iron to make steel. Material input-output relationships used to be important when transportation costs were relatively high. In the world of atoms, every city had a newspaper, and every paperboy had a route. Today, each country has only a few newspapers, and a few search engines are enough to serve the whole world. In the last decades the relative cost of moving knowledge, vis-à-vis the cost of moving the fruits of knowledge (products), has increased. Knowledge continues to be embodied in networks of people [12], and is concentrated in a few places, while bits and products travel swiftly around the world.

So why should we care about this "sticky knowledge principle?" On the one hand, new measures of relatedness are helping advance a more pragmatic industrial policy. The specificity of the principle is an antidote to the traditional temptations of industrial policy: the "cathedrals in the desert" built in naïve attempts to engineer development[1], or the gradualism that emerges when private sector companies capture public investment for sectors that are already well developed. Also, the principle is a remedy for the "everyone wants to be the new Silicon Valley" fever. But inoculation from erroneous ideas is not all that the principle provides.

What is sometimes counterintuitive, is that the principle of relatedness is not about over-specialization [13]. It is about understanding the unique paths that lead to diversification. Just like our understanding of gravity is linked to our dream to fly, our understanding of relatedness is linked to our desire to break the shackles of path dependency. The principle is linked to the desire of diversifying Chile away from Copper, or to have an Arab peninsula that is not solely reliant on exporting oil.

The principle also suggest that industrial policy should not be centered solely on identifying promising industries, but on identifying mechanisms that facilitate knowledge flows among industries and regions [14]. The policies supported by the principle are those focused on attracting the knowledge that regions are missing by

[1] Classic examples are the Volta River project in Ghana or the construction of a large Iron processing plant in Ipatinga Brazil.

facilitating the flow of the people who carry that knowledge [15], and by creating social bridges to the places where that knowledge is present. Some of these policies may involve traditional but important industrial policy instruments, such as industrial parks [16–18]. Other policies, focused on knowledge flows and collective learning, can sometimes suggest counterintuitive results. For instance, much of the world has human capital development policies that subsidize graduate studies while demanding the immediate return of students to their homeland. Yet, the bridges created by these scholars—and the respective “brain flows” they generate—can have a bigger impact in their homeland’s long-term collective learning than the repatriation of the human capital they absorbed in four years of college [19]. The principle of relatedness invites us to evaluate policies not based on short-term winners and looser, but on their ability to contribute to collective learning.

The principle of relatedness also warns us about some pitfalls. For instance, the principle of relatedness is a force that increases spatial inequality and can reduce the ability of peripheral cities to develop. That is, the principle of relatedness is better news for Boston than for Buffalo.

So where should this research move next? Exploring the channels that promote collective learning, unpacking the idea of relatedness, and identifying how and when countries and regions deviate from this principle are all fertile avenues of research. Recently, Pinheiro et al. showed that countries tend to deviate from this principle in about 7% of all cases, and that they are more likely to do so at an intermediate level of diversification [20]. Lee and Malerba, and Lee and Ki, showed that industry life-cycles provide windows of opportunities that can reshuffle industrial leaders [21, 22]. Boschma et al. showed that relatedness is a stronger force in countries with more coordinated forms of capitalism, than in countries with more liberal forms. Zhu et al. showed that policy interventions have been effective at moving Chinese regions into relatively unrelated areas [6]. Petralia et al. showed that technological relatedness matters more for developing than advance economies [23]. Murray et al. showed that strong intellectual property restrictions can reduce the number of new entrants in innovative activities [24]. And Uhlbach et al. [25] showed that in Europe R&D subsidies are more effective when they target areas with an intermediate level of relatedness, not too unrelated, but neither too closely related.

There are also theoretical questions, such as identifying the optimal diversification strategies that a country or region should follow. After all, recent work by Alshamsi et al. has shown that targeting the most related product is not always the optimal strategy, and that optimal diversification strategies need to worry about the right time to target relatively unrelated activities [26].

The good news is that learning appears to pay off for the economies who master it. Knowledge intense economies tend to grow faster [27, 28] and be less unequal [29] at the national level, than less knowledge intense economies, at the same levels of income, human capital, and similar institutions. Also, economies with a diverse set of related industries tend to experience faster employment growth in those sectors [30], and more entrepreneurship [32]. Workers in industries that enjoy the benefits of relatedness also are more robust to displacement shocks, since they find related work in their region more easily than comparable workers displaced in regions with few related activities [33]. These benefits tell us that efforts focused on promoting collective learning, among industries and regions, seem to be the right way to go.

For decades scholars have been trying to understand how cities, regions, and countries develop. The principle of relatedness does not provide a full answer to this question, but it signals a path. More importantly, the principle of relatedness is a beautiful illustration of the collaborative nature of science. Generalizing this principle to multiple scales, activities, institutional backgrounds, and metrics, has not been the work of a lone scholar, but the result of a diverse academic community involving geographers, economists, urban planners, physicists, and more. By building on each other's work, the scholars in this community have created a literature that supports the generalization of this principle. The reproducibility and robustness of this finding, therefore, does not rest on a single paper, instrumental variable, or key experiment, but on a growing literature that has replicated this principle tirelessly, demonstrating its solidity as a foundation for current and future research.

References

1. Marshall, A.: Principles of Economics. Macmillan and Company (1890)
2. Hidalgo, C.A., Klinger, B., Barabási, A.-L., Hausmann, R.: Science **317**, 482 (2007)
3. Neffke, F., Henning, M.: Strateg. Manag. J. **34**, 297 (2013)
4. Delgado, M., Porter, M.E., Stern, S.: J. Econ. Geogr. **16**, 1 (2016)
5. Neffke, F., Henning, M., Boschma, R.: Econ. Geogr. **87**, 237 (2011)
6. Zhu, S., He, C., Zhou, Y.: J. Econ. Geogr. **17**, 521 (2017)
7. Gao, J., Jun, B.A.: "Sandy" Pentland. Zhou, T., Hidalgo, C.A. arXiv:1703.01369 [Physics, q-Fin] (2017)
8. Kogler, D.F., Rigby, D.L., Tucker, I.: Eur. Plann. Stud. **21**, 1374 (2013)
9. Boschma, R., Balland, P.-A., Kogler, D.F.: Ind. Corp. Change **24**, 223 (2015)
10. Guevara, M.R., Hartmann, D., Aristarán, M., Mendoza, M., Hidalgo, C.A.: Scientometrics **109**, 1695 (2016)
11. Muneepeerakul, R., Lobo, J., Shutters, S.T., Goméz-Liévano, A., Qubbaj, M.R.: PLoS ONE **8**, e73676 (2013)
12. Hidalgo, C.: Why Information Grows: The Evolution of Order, from Atoms to Economies. Basic Books, New York (2015)
13. Glaeser, E.L., Kallal, H.D., Scheinkman, J.A., Shleifer, A.: J. Polit. Econ. **100**, 1126 (1992)
14. Balland, P.-A., Boschma, R., Crespo, J., Rigby, D.L.: Regional Studies 1 (2018)
15. Neffke, F., Hartog, M., Boschma, R., Henning, M.: Agents of structural change. The role of firms and entrepreneurs in regional diversification. Pap. Evol. Econ. Geogr. **14**(10). Utrecht University (2014). https://peeg.wordpress.com/
16. Zheng, S., Sun, W., Wu, J., Kahn, M.E.: J. Urban Econ. **100**, 80 (2017)
17. Chen, Z., Poncet, S., Xiong, R.: J. Comp. Econ. **45**, 809 (2017)
18. Kahn, M.E., Sun, W., Wu, J., Zheng, S.: The Revealed Preference of the Chinese Communist Party Leadership: Investing in Local Economic Development versus Rewarding Social Connections. National Bureau of Economic Research (2018)
19. Saxenian, A.: The New Argonauts: Regional Advantage in a Global Economy. Harvard University Press (2007)
20. Pinheiro, F.L., Alshamsi, A., Hartmann, D., Boschma, R., Hidalgo, C.A.: arXiv:1801.05352 [Physics, q-Fin] (2018)

21. Lee, K., Malerba, F.: Res. Policy **46**, 338 (2017)
22. Lee, K., Ki, J.: Res. Policy **46**, 365 (2017)
23. Petralia, S., Balland, P.-A., Morrison, A.: Res. Policy **46**, 956 (2017)
24. Murray, F., Aghion, P., Dewatripont, M., Kolev, J., Stern, S.: Am. Econ. J. Econ. Policy **8**, 212 (2016)
25. Uhlbach, W.H., Balland, P.A., Scherngell, T.: R&D policy and technological trajectories of regions: evidence from the EU framework programmes. Pap. Evol. Econ. Geogr. **17**(22), 1–21 (2017)
26. Alshamsi, A., Pinheiro, F.L., Hidalgo, C.A.: Nat. Commun. **9**, 1328 (2018)
27. Hidalgo, C.A., Hausmann, R.: PNAS **106**, 10570 (2009)
28. Hausmann, R., Hidalgo, C.A., Bustos, S., Coscia, M., Simoes, A., Yildirim, M.A.: The Atlas of Economic Complexity: Mapping Paths to Prosperity. MIT Press (2014)
29. Hartmann, D., Guevara, M.R., Jara-Figueroa, C., Aristarán, M., Hidalgo, C.A.: World Dev. **93**, 75 (2017)
30. Delgado, M., Porter, M.E., Stern, S.: Res. Policy **43**, 1785 (2014)
31. Frenken, K., Oort, F.V., Verburg, T.: Reg. Stud. **41**, 685 (2007)
32. Delgado, M., Porter, M.E., Stern, S.: J. Econ. Geogr. **10**, 495 (2010)
33. Neffke, F., Otto, A., Hidalgo, C.A.: Papers in Evolutionary Economic Geography **16.05** (2016)

Open Innovation in the Public Sector

Leonardo Ferreira de Oliveira[1,2](✉)
and Carlos Denner dos Santos Junior[1](✉)

[1] University of Brasília, Campus Universitário Darcy Ribeiro,
Brasília, DF 70910-900, Brazil
ldo235@nyu.edu
[2] Tandon School of Engineering, New York University,
6 MetroTech Center, Brooklyn, NY 11201, USA

Abstract. Innovation in the Public Sector has been seen as a way to respond the current challenges posed to society. However, the literature shows that little attention has been given to link public sector innovation to existing theories based on system and complexity. This essay aims to contribute to filling this gap with the development of a theoretical framework focused on the relation of organizational capabilities and public value generation in a systemic context. We discuss that the concept of Dynamic Capabilities - Sensing, Seizing and Transforming - is adherent to analyze public sector innovation, especially open innovation in government. In this context, the development of these organizational capabilities contributes to public value creation and can be analyzed through System Dynamics. We argue that these theoretical fields can be joined to elucidate complex and non-linear relationships related to open innovation in the public sector, making possible to scholars and policymakers to understand aspects of the systemic environment that can increase or decrease the creation of public value to providers, users, and beneficiaries.

Keywords: Public sector innovation · Open innovation · Public value
Dynamic capabilities · System Dynamics

1 Introduction

The concept of Open Innovation was developed by Chesbrough [1] regarding the private sector, meaning that organizations can search for advice or solutions through collaboration with external actors of the organization. The idea was transposed to the public sector given the several benefits that may come from open innovation adoption in government agencies, like awareness of social issues and benchmarking using citizen experience [2]. For instance, the Singapore's government announced in its 2018 budget that it is planning to develop an Open Innovation Platform where companies can list challenges that can be solved by digital endeavors.

Currently, the theme of Open Innovation in the public sector has called the attention of policy makers and scholars [3, 4], with calls for a research agenda [5]. This work focuses on the organizational dynamic capabilities [6, 7] present in public value creation [8–10]. This paper is organized in four sections including this introduction. The second section presents the concept of Public Value. Section three introduces the

A. J. Morales et al. (Eds.): ICCS 2018, SPCOM, pp. 458–466, 2018.
https://doi.org/10.1007/978-3-319-96661-8_47

theoretical model of Open Innovation in the Public Sector. The last section presents final considerations.

2 Public Value

Moore [11] explains that the purpose of public managers is to create public value according to three essential aspects. First, discern and formulate purposes that aim to create or improve public value. Second, gain legitimacy and support from its authorizing environment - considering stakeholders with different levels of legitimacy and power. Third, gain the capabilities needed to achieve goals of public value.

Several studies using Public Value Theory have been conducted [8, 12–15]. We follow the development provided by Harrison et al. [16], which draws attention that the rationale of public value concerns to the outcomes of government action related to the multiples possible types of public value. So, in this work, we understand Public Value Creation as "producing what is either valued by the public, is good for the public, including adding to the public sphere, or both, as assessed against various public value criteria" [17].

Another relevant aspect of Public Value is related to the perception of the effects of the organizational action. This perception might be diverse regarding the different roles several actors or groups may have, and also can variate in time. In this sense, government officials - acting as representatives of citizens - assess and prioritize the demand for service delivery through stakeholders' networks, creating public value [18]. Indeed, the public value notion also tackles a need for more recognition of the legitimacy of different stakeholders, as well as relies on its governance arrangements [14].

According to Freeman [19], stakeholders can be seen as "any group or individual who can affect or is affected by the achievement of the organization's objectives." Although the stakeholder view [19–21] was developed within a global vision of the organization, its use can be focused on specific situations in which the stakeholders take a position regarding an issue and are interested and can express a preference, such as in the case of corporate innovation projects [22, 23] or open data projects [24, 25].

This notion is important because often public managers find themselves challenged with a variety of stakeholders that might help them to accomplish the benefits of open innovation, but at the same time these stakeholders can also be seen as a risk if not correctly administrated [26]. For instance, Gonzales-Zapata and Heeks [27] explain how the different stakeholders and their perspectives on Open Government Data initiatives can be analyzed in a given context related with the nature, the drivers and benefits and the main actors of the initiative. In this way, using Open Government Data programs as a way to understand Open Innovation in the Public Sector, we follow the model developed for Dawes et al. [28]. The authors explain that stakeholders can be both internal or external to government and present a model that comprehend three main stakeholders groups:

(a) **Providers** – include government leaders and agencies in charge of Open Government Data initiatives - elected officials, managers and sometimes organizations that work as benchmarking in the field - promoting and pushing Open

Government Data initiatives forward. In synthesis, Political Leaders and Administrative agencies;

(b) **Users** – comprise Open Government Data users with the expertise to understand and analyze the open data, developing commercial and non-commercial applications. In synthesis, transparency advocates and Civic technology community; and

(c) **Beneficiaries** – include both individuals and organizations who adopt, buy, and use the products and services that Open Government Data initiatives have made possible. In synthesis, Consumers of open data products and services [28].

Based on Oliveira and Santos Jr. [29], we argue that the creation of Public Value can be a result of organizational capabilities, that will be perceived in several ways by different stakeholders at a specific moment in time, which in turn can start new pressures between the organization and the environment, resulting in predictable and in unpredictable changes in the public value delivered to society.

As said by Harrison et al. [16], conditions concerning public value "involve complex, non-linear relationships among diverse stakeholder groups and potential feedback loops," allowing us to bring to the discussion complex principles. In the next pages, we detailed a model (Fig. 1) with the use of the concepts of System Dynamics and present an analytical framework to the dynamics of open innovation in the public sector regarding public value creation.

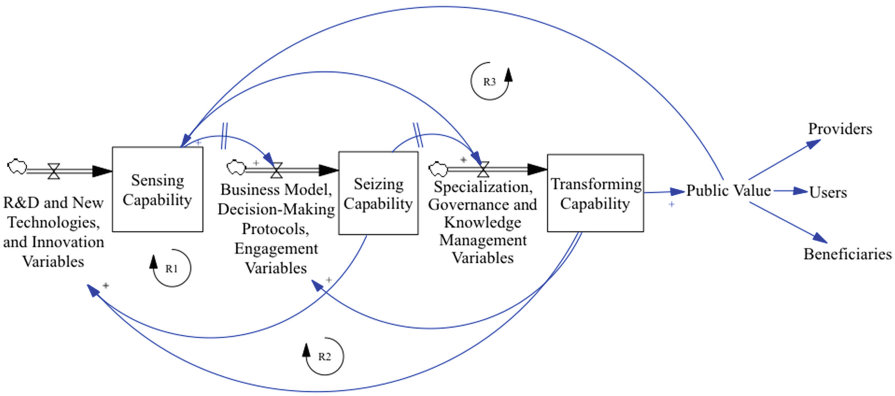


Fig. 1. Open innovation in the public sector, based on Oliveira and Santos Jr. [29].

3 Open Innovation in the Public Sector

Given the complexity of public sector innovation theme, some researchers indicate the use of a systemic approach to analyze it [29]. In a work that verifies what value does information technology generates in governments, Pang et al. [30] show that among resources and results there is a set of organizational capabilities which allow the transformation of inputs into outputs. We think that these capabilities can be adapted to demonstrate how public managers can generate public value through open innovation in the public sector. Following we described these capabilities.

3.1 Sensing Capability

In the explanation of the dynamic capabilities, Teece [31] comments that sensing (and shaping) is related to scanning, learning, and creating new opportunities. The author calls attention that two main classes of factors affect this capability. First, the access to existing information. Second, new knowledge or new information. This view is adherent to Moore's proposition related to the role public managers should have in the directions off its agencies - discern and formulate purposes with intent to create or improve public value [11].

Teece [31] explains that there are constraints regarding forces that exist in the environment, imposed by laws, regulators, standard bodies, ethics, etc. So, the "rules" that conduct the organizational action are the result of co-evolution and complex interaction between the ecosystems participants [31]. The same statement appears in the work of Rhodes [32], to whom the understanding of these "rules" might explain how public managers and the systems they act could be positioned to adapt to new circumstances and achieve improved outcomes through innovation.

Although the public value is pointed as a differential on innovations in the public sector compared to the private sector [33, 34], the limited literature on the verification of the results of changes and innovations in the public context calls attention [35, 36], since the main driver for public sector innovation is to create public value [37, 38].

We argue that the system dynamic view can be used as an approach to comprehend this phenomenon and realize the analytical system needed to sense and shape opportunities regarding the environment set. Thus, we used Sensing Capability as Stock Variable and based on the work of Teece [31] described its Flow Variables in:

- **R&D and New Technologies** – Processes to direct internal R&D and select new technologies from the environment set [31];
- **Innovation** – new ideas, processes and objects generated or developed with the intention to create public value [37, 39, 40].

For example, the *Portal de Datos Abiertos de la Justicia Argentina* (http://datos.jus.gob.ar/) is a judicial open data portal that aggregates 51 justice organizations in all country. The initiative is coordinated for the Ministry of Justice and Human Rights and is working to provide judicial statistics in open and re-utilizable formats. The action demonstrates that some agreements had to be done in order to make possible the innovation go through the courts. In the case, the Ministry of Justice and Human Rights assume a leading role to create environmental pressure that made possible to the justice organizations the development of new capabilities related to innovation.

3.2 Seizing Capability

Teece [31] explain that when a new opportunity is sensed, new products, services or process must be developed to address it – triggering strategic choice about technologies and resource allocation. The author also comments that addressing opportunities comprises preserving and improving technological competences to achieve acceptance and that the "design and performance specification of products, and the business model employed, all help define the manner by which the enterprise delivers value" [31].

We argue that the seizing capability is related with the legitimacy and support public managers have from its authorizing environment and also about the capabilities needed to achieve goals of public value, either within their own organization or by outside parties [11]. In this context, seizing is related to the business model public managers are conducting and how they communicate and receive feedback from several stakeholders. We used Seizing Capability as Stock Variable and based on the work of Teece [31] described its Flow Variables in:

- **Business Model** – Technology and product architecture; Targets; Mechanisms to capture value.
- **Decision-Making Protocols** – Inflexion points and complementarities; Risk aversion.
- **Engagement** – Asset specificity calibration; Stakeholders management; Leadership and communication; Culture. [31].

For instance, Dawes and Helbig [41] consider that transparency initiatives usually serve one of two goals. The first is related to offer citizens and other stakeholders a vision of what governments are doing and how it is functioning regarded the accountability of decisions and actions of elected officials and public agencies. The second is to present government data, making possible to generate social and economic value. In this respect, the Research and Documentation Centre of the Ministry of Security and Justice of Netherlands develops an initiative to share judicial data, improving transparency and developing collaboration and participation with other parties, and at the same time deals with the restrictions of the privacy protection principles and laws [42]. As Meijer, Conradie, and Choenni [42] explain that the Research and Documentation Centre developed a procedure to make decisions about open its data to the public (removing all privacy sensitive data), that under some pre-determined conditions permitting scientific organizations to have access to some privacy-sensitive research data. It was a new business model that provided new decision-making protocols and made possible the creation of engagement between agency and its stakeholders, leading way to public value generation.

3.3 Transforming Capability

To Teece [31] transforming is about the continuous alignment and reconfiguration of tangible and intangible assets, where success in the past will lead to the organization evolve in a path-dependent way. We argument that these characteristics can be seen as public manager action regarding the increase the quality or quantity of public activities per resource spent, and reduce the costs used to achieve current levels of production [11]. In this context, public organizations may tend to limit their investment in innovation that is well known and close to the existing asset base, as well as to minimize internal conflict and to maximize productive exchange inside the organization [31].

Another aspect this capability highlights is related to the Public Value Creation. As Meynhardt [43] explains the creation of Public Value is subjected to individual evaluation. So, in this work, we understand that the public service is delivered, but the Public Value is perceived for the different stakeholders. We used Transforming Capability as Stock Variable and based on the work of Teece [31] described its Flow Variables in:

- **Specialization** – Structures; Integration and coordination skills development; Strategic fit;
- **Governance** – Incentives; Agency issues; Strategic monitoring;
- **Knowledge Management** – Learning; Knowledge transfer; Know-how integration. [31].

For instance, the Chamber of Deputies of Brazil developed a project dedicated to open parliament practices. In this case, a new structure (the Hacker Lab) uses open data and collaborative tools as a way to access external knowledge and create engagement outside the government boundaries, promoting democratic discussions regarding the legislative policy agenda [44].

The Model (Fig. 1) depicts the existence of three reinforcement feedback loops. The First (R1) is related with the Sensing Capability influence in the Seizing Capability, meaning that the government action will start with the recognition of the new aspects (technologies and innovations) that will be shaped through the organizational action (business models, decision making and engagement). The second feedback loop (R2), is related to the action of transforming this new aspects, in interaction with the other two capabilities (Sensing and Seizing). The last feedback loop (R3), deals with the result of the interaction of the three capabilities (sensing, seizing and transforming) in a way to deliver public value to the stakeholders of the public organization.

4 Final Considerations

This article addressed the phenomenon of open innovation in the public sector, in order to understand its dynamics concerning public value creation. Much has been said about the role of innovation in the public sector [9, 10]. However, it is true that the current challenges imposed on the public administration reveal the need for a new attitude for both governments and society [29], including to contribute to the generation of innovations in government and public value to society.

We argued here that Sensing, Seizing and Transforming can be seen as organizational capabilities that might explain how open innovation in government occurs. As a next step, empirical research might assist in shaping the framework, as well as computer-based modeling may make possible a deeper comprehension of the phenomenon.

Acknowledgments. This work was supported by the Decanato de Pós-graduação – University of Brasília - DPG/UnB and by the Superior Court of Elections – Brazil.

References

1. Chesbrough, H.W.: Open Innovation : The New Imperative for Creating and Profiting from Technology. Harvard Business School Press, Boston (2003)
2. Mergel, I., Desouza, K.: Implementing open innovation in the public sector: the case of Challenge.gov. Public Adm. Rev. **73**, 882–890 (2013)
3. Szkuta, K., Pizzicannella, R., Osimo, D.: Collaborative approaches to public sector innovation: a scoping study. Telecommun. Policy **38**, 558–567 (2014)
4. Yun, J.J., Won, D., Hwang, B., Kang, J., Kim, D.: Analysing and simulating the effects of open innovation policies: application of the results to Cambodia. Sci. Public Policy **42** (2015). https://doi.org/10.1093/scipol/scu085
5. Kankanhalli, A., Zuiderwijk, A., Tayi, G.K.: Open innovation in the public sector: a research agenda. Gov. Inf. Q. **34**, 84–89 (2017)
6. Teece, D., Pisano, G., Shuen, A.: Dynamic capabilities and strategic management. Strateg. Manag. J. **18**, 509–533 (1997)
7. Wilden, R., Devinney, T.M., Dowling, G.R.: The architecture of dynamic capability research identifying the building blocks of a configurational approach. Acad. Manag. Ann. **10**, 997–1076 (2016)
8. Alford, J., O'Flynn, J.: Making sense of public value: concepts, critiques and emergent meanings. Int. J. Public Adm. **32**, 171–191 (2009)
9. Hartley, J.: Public value through innovation and improvement. In: Benington, J., Moore, M. (eds.) Public Value: Theory and Practice, pp. 171–184. Palgrave Macmillan, Basingstoke (2012)
10. Hartley, J., Alford, J., Knies, E., Douglas, S.: Towards an empirical research agenda for public value theory. Public Manag. Rev. 1–16 (2016). https://doi.org/10.1080/14719037.2016.1192166
11. Moore, M.: Creating Public Value: Strategic Management in Government. Harvard University Press, Boston (1995)
12. Benington, J.: From private choice to public value? In: Search public value – beyond private choice, pp. 1–36. Palgrave (2007)
13. Williams, I., Shearer, H.: Appraising public value: past, present and futures. Public Adm. **89**, 1367–1384 (2011)
14. Stoker, G.: Public value management: a new narrative for networked governance? Am. Rev. Public Adm. **36**, 41–57 (2006)
15. Alford, J., Hughes, O.: Public value pragmatism as the next phase of public management. Am. Rev. Public Adm. **38**, 130–148 (2008)
16. Harrison, T.M., et al.: Open government and e-government: democratic challenges from a public value perspective. Inf. Polity **17**, 83–97 (2012)
17. Bryson, J., Crosby, B., Bloomberg, L.: Public value governance: moving beyond traditional public administration and the new public management. Public Adm. Rev. **74**, 445–456 (2014)
18. De Tuya, M., Cook, M., Sutherland, M.K., Luna-Reyes, L.F.: Information requirements to create public value: sharing and opening data to address urban blight. Transform. Gov. People Process Policy **11**, 79–98 (2017)
19. Freeman, R.E.: Strategic management: a stakeholder approach, vol. 1 (1984)
20. Freeman, R.E., Reed, D.: Stockholders and stakeholders: a new perspective on corporate governance. Calif. Manag. Rev. **25**, 88–106 (1983)

21. Mitchell, R.K., Agle, B.R., Wood, D.J.: Toward a theory of stakeholder identification and salience: defining the principle of who and what really counts. Acad. Manag. **22**, 853–886 (1997)
22. Vos, J.F.J., Achterkamp, M.C.: Stakeholder identification in innovation projects: going beyond classification. Eur. J. Innov. Manag. **9**, 161–178 (2006)
23. Troshani, I., Doolin, B.: Innovation diffusion: a stakeholder and social network view. In: ECIS, vol. 10, pp. 176–200 (2006)
24. De Oliveira, L.F., Dos Santos Júnior, C.D.: The two sides of the innovation coin. In: 22nd Americas Conference on Information Systems (AMCIS 2016): Surfing the IT Innovation, pp. 1–19 (2016)
25. Zuiderwijk, A., Janssen, M.: The negative effects of open government data - investigating the dark side of open data. In: Proceedings of the 15th Annual International Conference on Digital Government Research - dg.o 2014, pp. 147–152. ACM Press (2014). https://doi.org/10.1145/2612733.2612761
26. Janssen, M., Charalabidis, Y., Zuiderwijk, A.: Benefits, adoption barriers and myths of open data and open government. Inf. Syst. Manag. **29**, 258–268 (2012)
27. Gonzalez-Zapata, F., Heeks, R.: The multiple meanings of open government data: understanding different stakeholders and their perspectives. Gov. Inf. Q. **32**, 441–452 (2015)
28. Dawes, S.S., Vidiasova, L., Parkhimovich, O.: Planning and designing open government data programs: an ecosystem approach. Gov. Inf. Q. **33**, 15–27 (2016)
29. Oliveira, L.F., Santos Jr., C.D.: Public value innovation: a theoretical framework based on system dynamics. In: Zuiderwijk, A., Hinnant, C.C. (eds.) Proceedings of the 19th Annual International Conference on Digital Government Research, dg.o 2018, Delft, Netherlands, May 30–June 1, 2018. ACM Press, New York (2018). Article 4, 9 pages. https://doi.org/10.1145/3209281.3209357
30. Pang, M.S., Lee, G., Delone, W.H.: In public sector organisations: a public-value management perspective. J. Inf. Technol. **29**, 187–205 (2014)
31. Teece, D. Explicating dynamic capabilities: the nature and microfoundations of (sustainable) enterprise performance. Strateg. Manag. J. **1350**, 1319–1350 (2007)
32. Rhodes, M.L.: Innovation in complex public service systems. In: Osborne, S.P., Brown, L. (eds.) Handbook of Innovation in Public Services. Edward Elgar Publishing (2013)
33. Hartley, J.: Innovation in governance and public services: past and present. Public Money Manag. **25**, 27–34 (2005)
34. Hartley, J., Sørensen, E., Torfing, J.: Collaborative innovation: a viable alternative to market competition. Public Adm. Rev. **73**, 821–830 (2013)
35. Kuipers, B.S., et al.: The management of change in public organizations: a literature review. Public Adm. **92**, 1–20 (2014)
36. De Vries, H., Bekkers, V., Tummers, L.: Innovation in the public sector: a systematic review and future research agenda. Public Adm. **94**, 146–166 (2016)
37. Moore, M., Hartley, J.: Innovations in governance. Public Manag. Rev. **10**, 3–20 (2008)
38. Kattel, R., et al.: Can we measure public sector innovation ? A literature review. LIPSE Work. Pap. 2–38 (2014)
39. Walker, R.M.: An empirical evaluation of innovation types and organizational and environmental characteristics: towards a configuration framework. J. Public Adm. Res. Theory **18**, 591–615 (2008)
40. Mulgan, G. Ready or not? Taking innovation in the public sector seriously, pp. 1–37. Nesta (2007)

41. Dawes, S.S., Helbig, N.: Information strategies for open government: challenges and prospects for deriving public value from government transparency. LNCS (Lect. Notes Comput. Sci. including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics), vol. 6228, pp. 50–60 (2010)
42. Meijer, R., Conradie, P., Choenni, S.: Reconciling contradictions of open data regarding transparency, privacy, security and trust. J. Theor. Appl. Electron. Commer. Res. **9**, 32–44 (2014)
43. Meynhardt, T.: Public value inside: what is public value creation? Int. J. Public Adm. **32**, 192–219 (2009)
44. Faria, C., Rehbein, M.: Open parliament policy applied to the Brazilian Chamber of Deputies. J. Legislative Stud. **22**(4), 559–578 (2016). https://doi.org/10.1080/13572334.2016.1235333

Trust Asymmetry

Percy Venegas(✉)

ECONOMY MONITOR, Heredia, Costa Rica
v.percy@economymonitor.com

Abstract. In the traditional financial sector, players profited from information asymmetries. In the blockchain financial system, they profit from trust asymmetries. Transactions are a flow, trust is a stock. Even if the information asymmetries across the medium of exchange are close to zero (as it is expected in a decentralized financial system), there exists a "trust imbalance" in the perimeter. This fluid dynamic follows Hayek's concept of monetary policy: "What we find is rather a continuum in which objects of various degrees of liquidity, or with values which can fluctuate independently of each other, shade into each other in the degree to which they function as money". Trust-enabling structures are derived using Evolutionary Computing and Topological Data Analysis; trust dynamics are rendered using Fields Finance and the modeling of mass and information flows of Forrester's System Dynamics methodology. Since the levels of trust are computed from the rates of information flows (attention and transactions), trust asymmetries might be viewed as a particular case of information asymmetries – albeit one in which hidden information can be accessed, of the sort that neither price nor on-chain data can provide. The key discovery is the existence of a "belief consensus" with trust metrics as the possible fundamental source of intrinsic value in digital assets. This research is relevant to policymakers, investors, and businesses operating in the real economy, who are looking to understand the structure and dynamics of digital asset-based financial systems. Its contributions are also applicable to any socio-technical system of value-based attention flows.

Keywords: Computational trust · Cryptocurrencies · Bitcoin
Behavioral finance · Web analytics · Blockchain analytics
Genetic programming · Markets disintermediation · Topological data analysis
Applied quantitative analysis · Fields finance

…reputation is a stock that is changed by the flow of good and bad actions… –Jay Forrester

The message is on the feedback –Gordon S. Brown

A. J. Morales et al. (Eds.): ICCS 2018, SPCOM, pp. 467–475, 2018.
https://doi.org/10.1007/978-3-319-96661-8_48

1 Introduction

A common concern among finance professionals, who usually make money by having access to privileged knowledge and special relationships, is how is it possible to do business when information asymmetries are close to zero—in shared distributed ledgers and blockchains, data is either public or available given proper authentication. By using examples from actual economic activity (in international trade and digital commerce) we can illustrate how the intuition of a *trust imbalance* may serve as starting point in the analysis. We define the concept of "trust asymmetry" in terms of dissimilarities in metric entropy (e.g. Kolmogorov Entropy) or as in this case, using symbolic regression complexity – which can be described in terms of the shape of a data space, and, the dynamics of vector fields.

Even if the information asymmetries across the medium of exchange are close to zero (as it is expected in a decentralized financial system), there exists a *trust imbalance*. And there are different levels of trust among trustful parties: naturally, a merchant will trust a local broker more than its foreign counterparty. Therefore, whoever can level-up trust provides a valuable financial intermediation service—at least until the system becomes mature. Ultimately, we will be able to answer questions such as: are the nodes running blockchain software essentially a material expression of people's beliefs? Particularly, is the "belief consensus" the fundamental source of intrinsic value that can be measured by intangible attention flows and tangible transactional activity?

2 Literature

An ideal scenario to study trust asymmetry is the case of a cryptocurrency fork, where at $t = 0$ one may assume equal conditions for the two chains (although in practice this is hardly the case, since the different fractions have already grouped around their preferred coin before the split, financial futures may have been trading already, and so on). In our paper on Crypto Economic Complexity [1], we argued that crypto economies tend to converge to the level of economic output that can be supported by the know-how that is embedded in their economy—and is manifested by attention flows. And, since a fork is really an event at the macroeconomic level (for instance, the economy of BitcoinCash vs the economy of Bitcoin), the aggregate demand for output is determined by the aggregate supply of output—there is a supply of attention *before* there is demand for attention. The socio-technical modeling of mass and information flow has usually been implemented in econometrics, industrial, and, policy planning circles, using Jay Forrester's System Dynamics methodology [2].

3 Methods

Data for this section includes digital assets historical monthly returns (Coincheckup.com), and daily time series for on-chain metrics (Coinmetrics.com), and, off-chain web and social analytics (EconomyMonitor.com and click-stream data providers). The period of study is August 2017 to January 2018.

3.1 The Characterization of Flows

When a blockchain split event occurs, a race (competition) for attention begins. Demand stars flowing from search engines, price trackers, faucets, wallets, educational sites, and the many services that support a crypto economy. One such event occurred on August 1st 2017, when the Bitcoin blockchain forked, creating two competing digital assets, BTC and BCH [3]. We obtained monthly data for 177 of those web services, specifically, the share of usage of each service towards two of the official communities in both networks; this off-chain activity acts as an inferential sensor, an indication of interest and trust. The sources were the largest contributors in the six-month period, their share is weighted by contribution. Due to its ability to identify and focus on driving variables, Symbolic Regression can build models from data sets that have more variables than records (these are commonly known as *fat arrays*, and most non-evolutionary machine learning techniques find issues to deal with them) [4]. Therefore, we apply genetic programming for dimensionality reduction purposes, and to build the predictive models that can provide insight into the shape of the *trust data space*.

The mechanism for generating input variables takes into consideration that the financial activity logged in the blockchain is the expression (e.g. a conversion event) of the consumption, flow of attention, and commitment of resources in enabling networks such as the web and mesh IoT. For instance, when a web browser creates a request (e.g. GET/HTTP) a Java-based application logs the hits, detects the device type (mobile or desktop), and other features included in the user-agent header. Also by looking down in the stack to the routing level, ISP data is used to obtain geographical origin, redirect path and destination.

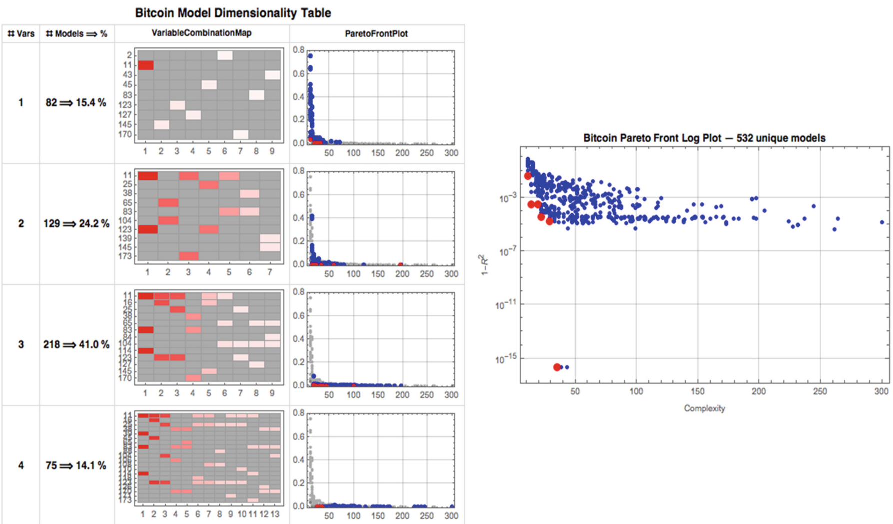


Fig. 1. Bitcoin, family of models (532 unique models). The period of study is August 2017 to January 2018.

Bitcoin. In total, 532 models were generated, with the majority of those (55.1%) containing at least three variables. The modeling process explores the trade-off between model complexity and model error (1-R^2). This is illustrated in the ParetoFrontLogPlot which displays each of the returned models' quality metrics, *complexit*y, and *accuracy*. The models denoted by red dots are all optimal in the sense that for a given level of accuracy there is no simpler model or, conversely, for a given level of complexity there is no more accurate model [5] (Fig. 1).

BitcoinCash. In the BitcoinCash case it is more difficult to determine what the dominant best models are, this is confirmed by the number of models with relatively high error, and the higher dimension and larger number of possible variable combinations (50 models use 5 variables, wherein the Bitcoin case no model had more than 4 variables).

4 Analysis

What we would like to understand from the *shape of the data* is what are those factors which variability has a noticeable impact on actual investor expectations changes, as measured by *price returns*–even if those sources are not among the largest traffic contributors to the crypto economy. We would also prefer to focus on the models in the knee of the Pareto front since those represent the better trade-off between complexity and accuracy.

For Bitcoin, the model with complexity equal to 22 becomes informative. It contains a metavariable (*laser.online * vKontakte*) that appears in 6.6% of the models, and one of the variables from that specific metavariable construct (*laser.online*) appears in some form in 4 of the 6 finalist models. This is notable because while the other two variables in the model are a proxy for demand (the largest social network and search engine in Russia), usage of laser.online actually has investment implications – that service was a famous bitcoin scam and Ponzi scheme, where BTC holders actually invested and lost funds [6]. The p-value for the metavariables considered in the analysis is under 0.03, as shown in Fig. 2.

In the BitcoinCash economy, the drivers are notably different, and the complexity of the models tends to be higher. The fact that the independent variables are different than those that drive BTC returns speaks for the structurally different constitution of both economies: users of different services both consume investment information and have a preference to trade in different exchanges, such as Korea-based Bithumb. Some are even different people, as demonstrated by the fact that they seek information in Yahoo and social validation in Facebook, not in Yandex and vKontakte. The higher information content of higher complexity models may also induce over-fitting in the presence of noise.

What is notable is that the flows are diverse in terms of data sources, with everything from due diligence resources to entertainment—in other words, the idea that cryptocurrency market formation is only fueled by financial speculation is misleading. Each of these economies is a living organism that has a distinctive evolution that can be measured from the inception, or, from the time of the fork. And as complex organisms,

Bitcoin Model Selection Report

	Complexity	$1-R^2$	Function
1	11	0.040	$4.80 - \frac{2.90\times10^{-2}}{\text{duckDuckGo}}$
2	14	3.155×10^{-4}	$-0.26 + 368274.40\ \text{laser•online}\ \text{vKontakte}$
3	19	2.836×10^{-4}	$-0.32 + \frac{5.50\times10^{-6}}{\text{downloadhelper•net}} + 279.78\ \text{laser•online}$
4	22	3.543×10^{-5}	$-0.20 + 369048.04\ \text{laser•online}\ \text{vKontakte} - \frac{1.06\times10^{-3}}{\text{yandex•Search}}$
5	29	1.594×10^{-5}	$-0.25 - (2.28\times10^{8})\ \text{bitinfocharts•com}\ \text{criptonoticias•com}^2 + 367919.50\ \text{laser•online}\ \text{vKontakte}$
6	35	0.000	$-0.85 - 4700.09\ \text{bitcoinpaw•com} + 819.39\ \text{coindesk•com} - 3551.32\ \text{coinranking•com} + 2399.03\ \text{hongkiat•com}$

Meta Variable Table (Significance level 0.03)

Rank	Count	MetaVariable	% of models	% of MetaVariables
1	78	$\frac{1}{\text{duckDuckGo}}$	14.7	17.0
2	52	$\frac{\text{bitcoin•pl}^{6.}}{\text{coinranking•com}^{6.}}$	9.8	11.4
3	52	$\frac{1}{\text{coinranking•com}^{6.}}$	9.8	11.4
4	52	$\left(\frac{\text{bitcoin•pl}^{6.}}{\text{coinranking•com}^{6.}}\right)^{1/3}$	9.8	11.4
5	52	$\text{bitcoin•pl}^{6.}$	9.8	11.4
6	51	webmoney•ru^2	9.6	11.1
7	49	$\frac{1}{\text{webmoney•ru}^2}$	9.2	10.7
8	49	$\text{laser•online} + \text{linkedin}$	9.2	10.7
9	49	$\frac{1}{\text{laser•online}+\text{linkedin}}$	9.2	10.7
10	47	$\frac{1}{\left(\frac{1}{\text{webmoney•ru}^2}\right)^{0.125}}$	8.8	10.3
11	39	$\text{duckDuckGo}^{1/3}$	7.3	8.5
12	38	duckDuckGo^2	7.1	8.3
13	37	$-\text{coinranking•com} + \text{duckDuckGo} - \text{hongkiat•com}$	7.0	8.1
14	35	$\text{laser•online}\ \text{vKontakte}$	6.6	7.6
15	33	$-2.\ \text{coinranking•com} + \text{duckDuckGo} - \text{hongkiat•com}$	6.2	7.2
16	33	duckDuckGo^3	6.2	7.2
17	26	$\sqrt{-\text{coinranking•com} + \text{duckDuckGo} - \text{hongkiat•com}}$	4.9	5.7
18	20	$\text{duckDuckGo} + \text{laser•online}$	3.8	4.4

Fig. 2. Bitcoin returns as a function of off-chain activity, selected models (significance level 0.03). August 2017 to January 2018.

they may have different levels of viability depending on the connectedness, influential actors, and risk present on their associated networks.

4.1 Trust Asymmetry: The Quantitative Approach

Shape. The comparison of the development of symbolic regression expressions over generations provides the first proof of dissimilarity between the crypto economies; consistent with the nascent stage of the BitcoinCash economy, twice as many generations are required to model the returns as a function of inflows when compared to the more mature Bitcoin economy (see Fig. 3).

Furthermore, it is possible to map the asymmetry of trust using a combination of multidimensional scaling, a statistical technique, and topological data analysis [7] a novel type of econometric analysis which complements the standard statistical measures and has been used to detect early warning signals of imminent market crashes.

We begin by selecting functions from the Model Selection Report with complexity at the same accuracy level (e.g. 10^{-4}). Secondly, we draw a graph where a set of vertices ($v_1, \ldots, v_N$) is an element of V connected by M edges ($e_1, \ldots, e_M$) that are elements of E, where the length of each edge, ($l_1, \ldots, l_M$) are elements of L. The edges mirror complexity values, giving rise to a complexity space (in our case, a *trust space*, since at least one model included in the subset contains a variable that is a direct expression of investor's financial commitment – such as the use of a cryptocurrency exchange). Figure 3 (bottom) shows a tangible representation of the trust imbalance concept. By comparing the edge lengths (Euclidean distance) and complexity values using a ratio of the form distance/complexity, we find that the median distance is 0.00235542 for Bitcoin and 0.000860686 for BitcoinCash. The counterintuitive finding is that although the Bitcoin economy is more complex in macroeconomic complexity terms (diversity, and ubiquity of services), during the stage of formation of the competing BitcoinCash economy the complexity of models required to describe it is higher,

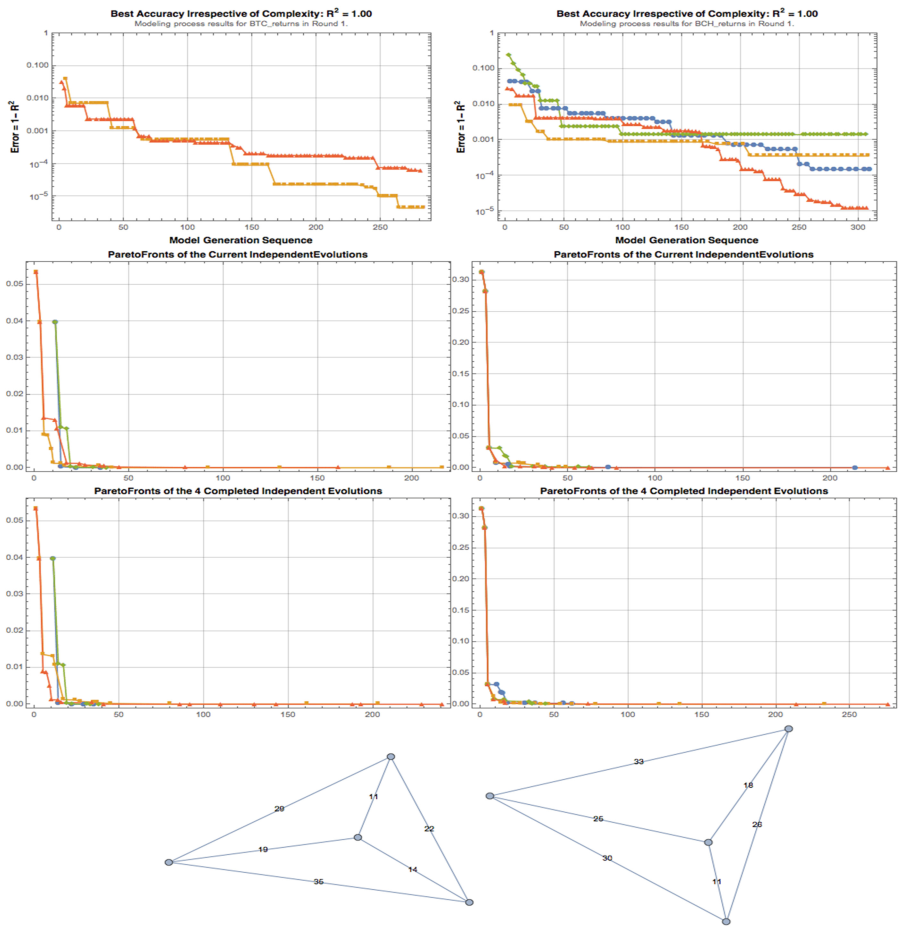


Fig. 3. Evolutions (top). Shapes of the complexity space (bottom). Left: Bitcoin. Right: BitcoinCash.). August 2017 to January 2018.

given a similar level of accuracy. That is, even in terms of structure, the older economy is in a relatively steady state in relation to the new entity.

Dynamics. To model the dynamics we make use of Forrester's System Dynamics approach, a tool familiar to econometricians and policymakers. If we simplify to obtain the form of a two-sided system (what one economy loses the other gains) and focus on the flows in one direction, the schematic is as shown in Fig. 4 (left). The "goal" is an implicit input to the top component, and the flow of attention (with a gauge that implies a variable rate of action) is an input to the stock component at the bottom; the feedback loop represents information about the state of the level of trust.

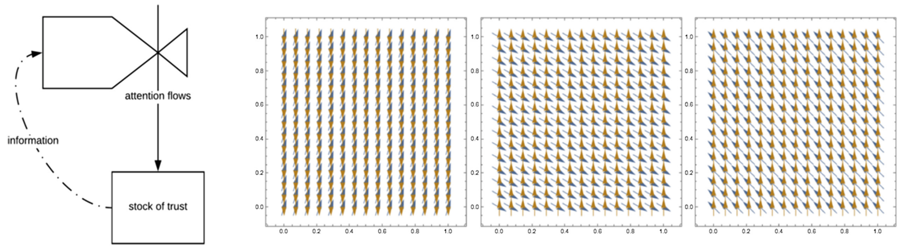


Fig. 4. Level of Trust (left). Gradient field over the unit square (right). September, October and November 2017 (usual order).

In analytical form, the general equation that describes the stock component is

$$level = \int_{n}^{m} \Sigma(in\_flows) - \Sigma(out\_flows) \tag{1}$$

Where n, m denote the complexity boundaries; we integrate over time, since we are measuring usage per month. The outflows are implied, and not shown in the graph, but we assume that whatever attention BitcoinCash is losing, Bitcoin is gaining –although in practice there might be as well leakages towards other competing forks.

As a case in point, at complexity level 11 (the worst error is what matters) BCH returns are driven by inflows into the BitcoinCash economy.

$$BitcoinCash_1 = \left(-0.13 + \frac{4.27 \cdot 10^{-6}}{github}\right). \tag{2}$$

And outflows can be described by the Bitcoin economy gains.

$$Bitcoin_1 = \left(4.80 + \frac{2.90 \cdot 10^{-2}}{duckDuckGo}\right). \tag{3}$$

At the same level of complexity the error measure associated to the Bitcoin model (0.04) is lower than for the BitcoinCash model (0.283); again, the result demonstrates the behavioral traits of the economic agents, as you would expect attention flows towards a software code repository (Github) become a factor for the newer coin, while the more established coin has higher visibility in organic channels (in this case, duckDuckGo, a search engine popular among developers). This formulation encapsulates tacit knowledge since the model includes information in people's heads (e.g. search patterns are revealed preferences, but are private to the user until the data is mined).

Fields Finance. Another way to analyze the condition of asymmetry is by looking at trust imbalances among the same set of variables. In this manner, we force the evolutionary algorithm to choose the best model that simultaneously contains both variables, and that allows for the flow to be visualized on a higher dimensional space (e.g. a vector field).

To make the streams fully descriptive of the path to material economic activity (not simply market sentiment) we use blockchain fees rather than returns, and time series of daily usage data rather than share of inflows; the off-chain data expressively includes variables related to transactional activity (e.g. cryptocurrency exchanges, cryptocurrency payment platform for merchants). This allows for a better description of the causal relationship, and facilitates additional verification using forecasting methods such as bivariate Granger causality [8].

The resulting inflow equations are arranged into a field of the form given by

$$\{Fees_{BCT}, Fees_{BCH}\} = \{f(X_1, X_2), g(X_1, X_2)\}. \quad (4)$$

Where X_1 refers to huobi.pro, a Chinese exchange; X_2 refers to coinpayments.net, a payments platform.

We slice the data by month (from September to November), to focus on the periods of analysis that are of interest – where we want to study the persistence or the break of trust symmetry. We obtain 6 equations in total, 2 for each month (each one describes how usage of the services under study may predict the movements in BTC or BCH fees). To obtain the rate of change of inflows levels (rather than levels themselves) we use the expression (1) applying a derivative at both sides and without other modification than assuming outflows equal 0; this requires that we compute the gradient of the field. The results are plotted in Fig. 4 (right), where X_1 is the component in the horizontal axis. The flows give rise to a field. The *f* term (blue) and the *g* term (brown), each is expressed by vectors with components X_1 and X_2. We see how in September the vectors almost cancel out each other, however, the effect of consumption on BTC is leading –in fact, that month neither the exchange or the shopping cart solution show flows that are meaningfully correlated to the BitcoinCash economy. The economy of BitcoinCash actually becomes relevant to these services in October; as for Bitcoin, fees behavior is better described by the rate of exchange usage in October and by the rate of merchant service usage in November. The flows are mapping the belief consensus of the users of each coin. The transition from September to October marks the phase change in trust dynamics (when the new coin adoption actually kicks off among the general public).

5 Conclusions

Digital assets detractors usually say that there is no proven demand for cryptocurrencies, but it has been demonstrated that demand not only can be measured but that crypto-economies and their driving variables can be ranked as demand evolves [1]. And since in blockchains transaction count and exchange volume can be manipulated by batching transactions and other artifacts [8], one of the viable measures of value might be actual supply and demand of attention. Furthermore, if crypto assets defy the "Efficient Market Hypothesis" and the idea that all available information is encoded in prices, something more profound may be going on here: beyond any of the traditional definitions of utility, *disintermediation of trust* by itself might entail a premium. In that case, the value of the chain may reside on the chain itself: the nodes running the software are simply an expression of people's beliefs—being that the belief that the market can be manipulated for personal gain; that it is about time to challenge the government monopoly on money; that algorithmic money might be the more convenient utilitarian artifact to conduct transactions if you have already digitized a large part of your day-to-day activities; or else. This belief consensus is a human-machine construct, and perhaps this is why economists who are not trained as technologists have a hard time grasping the implications of a blockchain financial system. And, a more fundamental question about value arises: as the trust asymmetries between crypto economies reveal a structural divergence in value perception, could this paradigm provide incontestable proof of value in digital assets, including those with enhanced privacy features which by default make key transactional data opaque or unavailable?

References

1. Venegas, P.: Crypto Economy Complexity. Authorea Inc. (2017). https://doi.org/10.22541/au.151979285.51792474
2. Forrester, J.W.: System dynamics—the next fifty years. Syst. Dyn. Rev. **23**(2–3), 359–370 (2007). https://doi.org/10.1002/sdr.381
3. Bitcoin Cash: Wikipedia (N.d.). Accessed 21 Feb 2018
4. Riolo, R., Moore, J.H., Kotanchek, M. (eds.): Genetic Programming Theory and Practice. Springer International Publishing (2016). https://doi.org/10.1007/978-3-319-34223-8
5. Kotanchek, M., Smits, G., Vladislavleva, E.: Pursuing the pareto paradigm tournaments algorithm variations and ordinal optimization. In: Genetic Programming Theory and Practice, pp. 167–185. Springer (2007). https://doi.org/10.1007/978-0-387-49650-4_11
6. Reddit. https://www.reddit.com/r/Bitconnect/comments/7i2wns/laseronline_website_address_reroutes_to_bitconnect/. Accessed 21 Feb 2018
7. Gidea, M., Katz, Y.: Topological data analysis of financial time series landscapes of crashes. Phys. A Stat. Mech. Appl. **491**, 820–834 (2018). https://doi.org/10.1016/j.physa.2017.09.028
8. Seth, A.: Granger Causality. Scholarpedia **2**(7) (2007). https://doi.org/10.4249/scholarpedia.1667
9. Coinmetrics homepage. https://coinmetrics.io/difficulty-estimating-chain-transaction-volume/, https://coinmetrics.io/difficulty-estimating-chain-transaction-volume/. Accessed 21 Feb 2018

Exploring the True Relationship Among Countries from Flow Data of International Trade and Migration

Kedan Wang[1,2], Xiaomeng Li[1], Xi Wang[2], Qinghua Chen[1(✉)], and Jianzhang Bao[1(✉)]

[1] School of Systems Science, Beijing Normal University, Beijing 100875, China
{qinghuachen,baojz}@bnu.edu.cn

[2] School of Government, Beijing Normal University, Beijing 100875, China

Abstract. The relationship among various entities in the socio-economic systems is an important part of complexity research. Here we combine the general gravity model and minimum reverse flows idea to propose a general framework to reveal comprehensive relationship among entities with intimacy and hierarchy based on flow data among entities. Besides, we apply this method to comprehensively analyze international trade network and population migration network. Based on the empirical flow data, we calculate the effective distance among countries and rank or grade of countries, which could reveal the true relationship among them. The countries in global trade are clustered but not hierarchical, while the relationship among countries in international migration is just the opposite. They are hierarchical and not clustered.

Keywords: Relationship among countries · General gravity model
Minimum reverse flow · International trade · International migration

1 Introduction

The relationship between various factors is an meaningful feature for complex socio-economic systems. To reveal the relationship of factors based on real data is a interesting issue, attracting researchers' attention. For example, the causality relationship of energy consumption and GDP [1], the relationship between stock prices and exchange rates [2], economic growth and economic development [3], financial development and economic growth [4]. Comparing to the connections and roles between detail factors, the relationship among integrating entities is neglected often. And this kind of comprehensive relationship, such as the relationship in between countries, is especially important for discovering some of the world patterns [5].

It is significantly useful to reveal the true relationship among countries in order to understand the global economic, politics and culture. Besides that, based on reliable analysis, countries' development trends can be predicted, which

A. J. Morales et al. (Eds.): ICCS 2018, SPCOM, pp. 476–485, 2018.
https://doi.org/10.1007/978-3-319-96661-8_49

can provide advice and suggestions to government agencies for making proper policy decisions. However, in the socio-economic systems, all countries in the world are linked, and they connect and intricately interact with each other. The relationship between countries is complex and has strong dynamic characteristics which are hard to grasp.

In recent years, certain researchers proposed new skills to map the countries into the appropriate space, and found some methods to describe the countries' relationship more appropriately and intuitively. As a representative study, Hidalgo et al. build a so-called product space, and each country was being represented by the spatial point cloud determined by the exported product [6]; Guillermo et al. projected the countries directly at a point in a hyperbolic space by international trade statistical data with a kind of gravity model [7]. The essence of these works is trying to understand the relationship in between countries in worldwide trade.

These works show the relationship among countries visually through actual data that has a very strong objectivity, but they are not comprehensive enough to make further analysis and application successfully. More specifically, for example, they hardly explain the factual existence of inequality in global trade and migration at all. To make up for this shortage, this paper proposes a new framework for adding analysis of hierarchical relationships to the discussion of alienation relations. A systematic method for understanding the true relationship between nations will be proposed and applied in the international trade and the international migration. The paper is organized as follows: In Sect. 2, we introduce the data source and methods used in this paper. Section 3 exhibits the results of its application for the global trade and the global migration. In the last Section, the conclusions and discussions are provided.

2 Model and Data Sources

The object we analyze in this paper is a complex network. The nodes of the network are countries, and the edges are the flows between countries, which can be trade flow, or immigration flow, or others. We use a general gravity model to find the effective distances between nations and use the method of minimum reverse flow to discuss hierarchical relations of the countries.

2.1 General Gravity Model

Newton's Gravity Model has not only been successfully cited in explaining the interaction of two arbitrary objects in physics, but has also been successfully applied in the analysis of certain socio-economical systems, such as commuting [8] and telecommunication flows [9] between two cities, even trade [10] and migration [11] in between countries. In this paper, we apply a general gravity model to describe the influence of factors to both trade and migration flows among countries as follows.

$$F_{i,j} = G\frac{(M_i \times M_j)^{\alpha}}{D_{i,j}} \tag{1}$$

Here, $F_{i,j}$ is the flow from country i to j. M_i and M_j are data of an attribute of country i and j. It is GDP for the global trade but total population for international migration. G and α are constant exponents. $D_{i,j}$ is the effective distance between i and j and there is $D_{i,j} = D_{j,i}$. This kernel concept effective distance reflects the degree of intimacy and alienation between countries. Countries with smaller effective distance are more likely to emigrate and international trade. On the contrary, the greater the effective distance between countries, the less fluent their immigration and international trade channels are.

In its alternative form using logarithms, we obtain,

$$logF_{i,j} = logG + \alpha log(M_i \times M_j) - logD_{i,j} \tag{2}$$

Then, we use the Least Square Method to find the optimum parameters with Eq. 3.

$$\min_{G,\alpha,D_{i,j}} \sum \|F_{i,j} - (logG + \alpha log(M_i \times M_j) - logD_{i,j})\|^2 \tag{3}$$

2.2 Minimum Reverse Flow

The sorting or ranking can represent the hierarchy of elements in a system and there are different sorting methods for different systems [12]. Recently, the idea of minimum violation ranking based on the complex network provides a general algorithm which can be widely applied in various fields. Clauset successfully used that for ranking universities based on the flows of doctorate graduates [13]. We adopt that minimum violation ranking to minimum reverse flows. The idea is described in Eq. 4.

$$\min_{...\pi(i),...,\pi(j)...} S(A) = \sum_{i,j} F_{i,j} \times sign[\pi(i) - \pi(j)] \tag{4}$$

There $S(A)$ is the score of the system in sort of A under condition $...\pi(i),...,\pi(j)...$. $F_{i,j}$ is the flow from country i to j. The ranking of country i and j are $\pi(i)$ and $\pi(j)$ respectively. $sign()$ is the sign function. If the order of i is in front of j, that is, the value of $\pi(i)$ is less than the value of $\pi(j)$, then this $sign[\pi(i) - \pi(j)]$ is negative, and the contribution $F_{i,j} \times sign[\pi(j) - \pi(i)$ is positive. On the contrary. If the order of j is in front of i, the contribution will be negative. The idea of minimum reverse flow means to find the optimum solve A^* to the minimize $S[\pi(A)]$ in any possible order A.

A schematic diagram in Fig. 1 can be described more intuitively. The original network structure will be stretched to a sequence in a case of ranking. All flows are displayed in the left and right direction. If the order from left to right is a value from least π to biggest one, then the sum of the flows to the left is required to be the smallest.

We use the stochastic optimization technique and bootstrap method mentioned in the reference [13] to get the π values of different countries, thus determining the country's hierarchical position. There is a very simple index ρ to measure the strength of system hierarchy. The ρ represents the fraction of flow

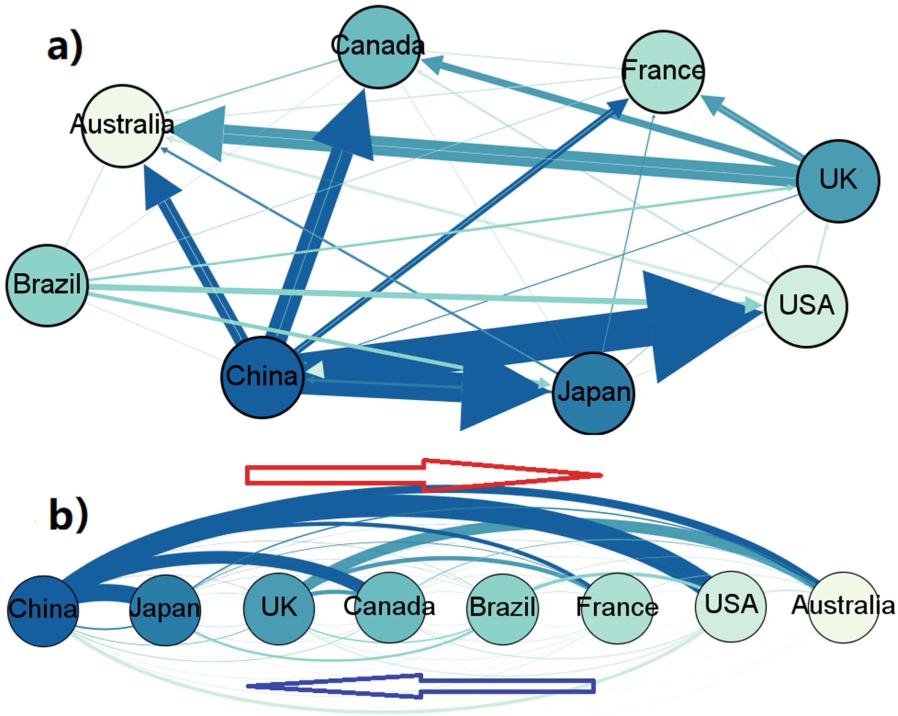


Fig. 1. Schematic diagram of minimum reverse flow. (a) The flow networks of 8 sampling countries. Each edge width represents the flow amount from country i to country j. (b) Optimum order for these 8 countries. From left to right, the ranking number of the countries increases. The reverse flows which from right nodes to left ones are shown below the nodes. Minimum Reverse Flow requires to minimize the amount of that part flows.

$F_{i,j}$ pointing downward. ρ will get its maximum value under the request of minimum reverse flows. When $\rho \to 1/2$, countries have no significant hierarchy. But $\rho \to 1$ indicates a serious hierarchy.

2.3 Data Sources

The whole description of the data source used in this study is given in Table 1. The global trade data is from the International Trade Statistics Database of UN Comtrade. According to the integrity and reliability of the data, we have chosen the trade situation of 148 countries/regions. The cross-country trade data of 148 countries or regions is downloaded and organized. Besides that, the data of the GDP of these countries are got from the World Bank database.

The cross-country migration data used in this paper is got from World Bank. Based on the availability and integrity of the data, we select 153 countries/regions with the classification. The differences between the bilateral

Table 1. The data source

Indicator	Indicator description	Data source
Trade flow between countries	Trade statistics (current US$), 2016	https://comtrade.un.org/
GDP	GDP (current US$), 2016	https://data.worldbank.org/indicator/ny.gdp.mktp.cd
Migrant flows for other countries or regions except China	Estimates of migrant stocks 2000, 2010	http://www.worldbank.org/en/topic/migrationremittancesdiasporaissues/brief/migration-remittances-data
Migrant flows for China	For China only has total Migrant Stocks in world bank's database. China's bilateral flow data is from the 2010 Population Census of The People's Republic of China	http://www.stats.gov.cn/
Population	Population, 2000–2010	https://data.worldbank.org/indicator/sp.pop.totl

estimates of migrant stocks in 2000 and 2010 describe the number of migration flows. Especially for China's data, we obtained it from the "2010 Population Census of The People's Republic of China".

3 Results

3.1 The Fitting of Gravity Model

We fitted the flows of the global trade and migration data with Eq. 3. The function could describe both the trade and migration flows very well, and the R^2 statistic is 0.67 for 2016 global trade network and 0.68 for 2000–2010 global migration network. The Fig. 2 shows that the estimated flows are very close to the original data respectively.

3.2 Location of Countries in the Global Trade and Migration Spaces

Here with the result of the effective distance matrix, we could locate the countries in the two-dimensional space with MDS method which projects objects into 2-dimensional space [14]. The Euclidean distance between countries approximately expressed the effective distance.

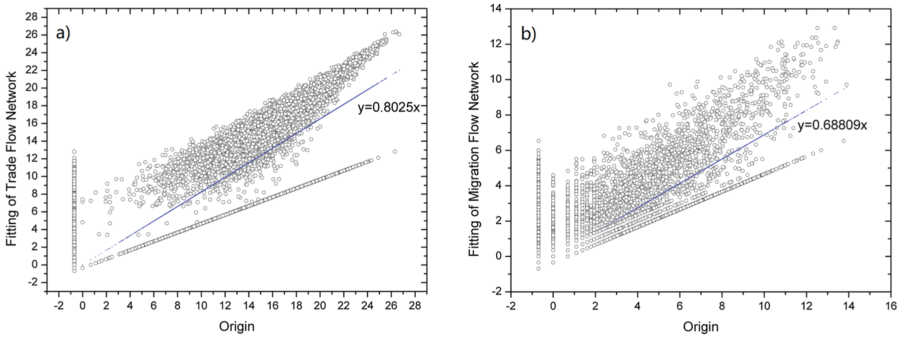


Fig. 2. Fitting results for global trade and migration flows (by double logarithm scale). (a) is for global trade and (b) is for global migration. The horizontal axis represents the original flow data, and the longitudinal axis represents estimated flow amount by Eq. 3. The R^2 statistic is 0.67 for 2016 global trade data and 0.68 for 2000–2010 global migration data. The blue lines are regressions to the scatter points.

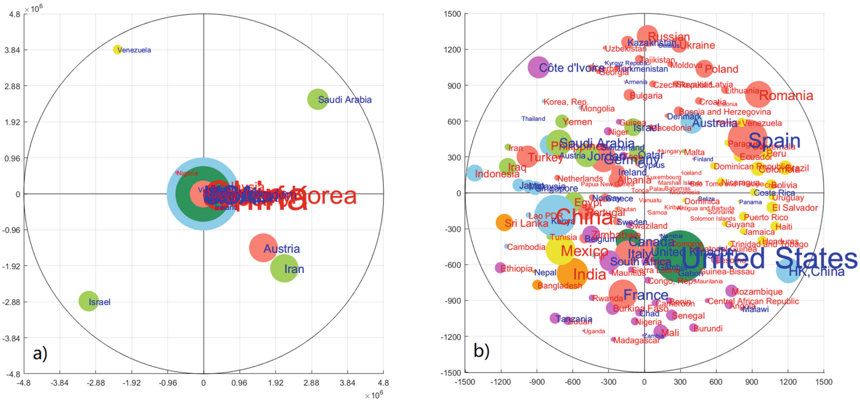
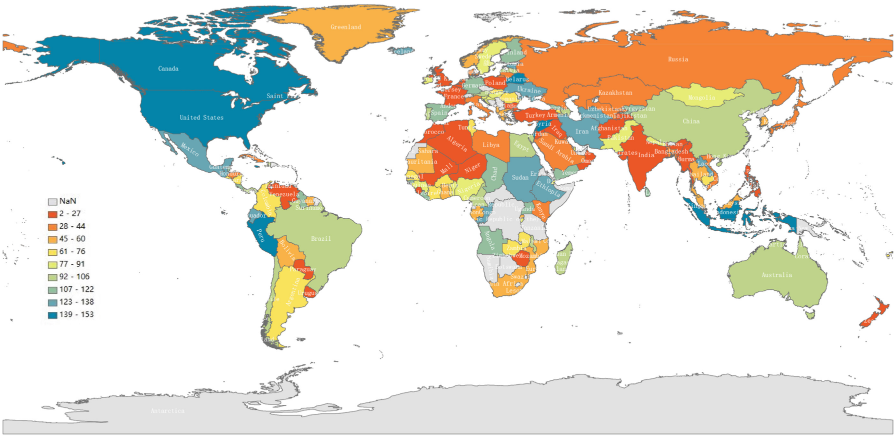
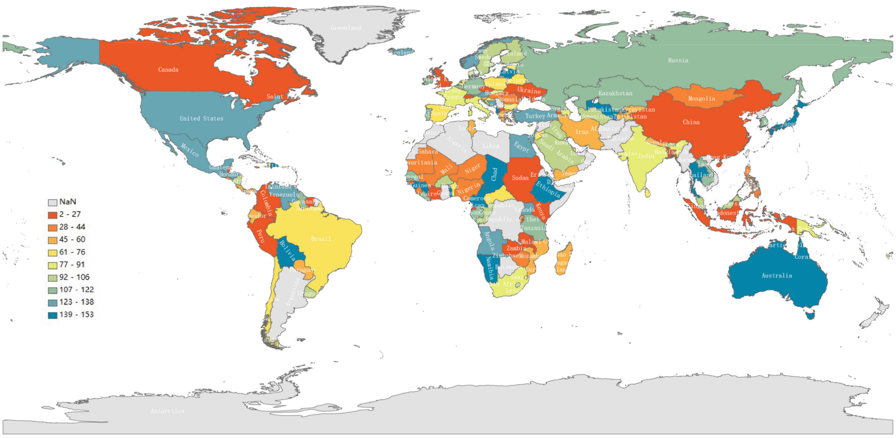


Fig. 3. Approximate locations of countries in the flow network. (a) is for global trade and (b) is for global migration. The embedding results get from effective distances between countries by MDS method which projects data from high dimensional space to the two-dimensional. The closer the two points are, the more effective distance the corresponding countries have. The color of countries' name means the direction of trade/immigration flow, in which red means net outflow and blue means inflow; the color of dots represents the geographical location of the country; and the size of dots describes the amount of net flow.

Figure 3 shows the location of countries in global trade network and migration network. It is obvious that there is more imbalance in the trade network, where more countries are clustering. Except that, the position of the countries can show certain relationships clearly. The countries locating in the center position have small average effective distances with the other countries. Conversely, the



(a) Rank map for countries in global trade network



(b) Rank map for countries in global migration network

Fig. 4. The hierarchy of countries under the international trade and migration situations. The color for each country demonstrates their rankings. If countries have higher rankings, their colors in the map will be closer to blue, and if countries have lower rankings, their colors will be closer to red. The yellow ones rank middle. The countries without available data are in gray.

countries on the edge are isolated in the global flow situation. The distance between the countries could be used as a parameter to predict the two countries are likely to have trade/population contacts.

In detail, France is the country with the smallest average effective distance to other countries. Italy, Turkey, Germany, China, United Kingdom, and The United States have relatively small average distance (subplot a) of Fig. 3). In general, these countries are relatively easier to trade with the other countries in the world. Among the population migration optimization results, Australia has the smallest average distance to other countries (subplot b) of Fig. 3). However, from the embedding result, the world pattern of population migration is uniform and decentralized.

3.3 Hierarchy of Countries in Global Trade and Migration Network

Figure 4 shows the hierarchy of countries. The color of each country demonstrates their rankings. Countries are ranked smaller if the colors are redder. Countries are ranked bigger if the colors are bluer. The yellow ones rank middle. The countries without available data are in gray.

To some degree, the ranking indicates the net in-flow or out-flow. In migration networks, for example, USA ranks 132 (Fig. 4b). The immigration to USA is much higher than the emigration from USA, while the situation is reversed in China and Colombia, which rank 5 and 9.

The value of ρ in trade flow networks is 0.65, which indicates that the international trade networks do not have a strong hierarchy. However, in migration networks, $\rho = 0.92$. It shows that there is a strict hierarchy. This result indicates that the factors like migration policy and willingness of immigrants have effects on the choices of immigration destination, which forms a ranking where migrants are not likely to immigrate to countries with higher ranks.

4 Conclusions and Discussions

In this paper, a comprehensive and systematic framework for discussing the comprehensive relationship among nations is put forward. We combine a general gravity model to find the relationship in between nations and use the method of minimum reverse flow to discuss hierarchical relations in between countries in case of international trade and immigration.

Firstly, we find the general gravity model can fit both the international trade flow data and migration data very well. Compared with other research in relative literature as [7] where the entity relationship in the network is considered just connected or not, the advantage of this analytical method is that all accurate flow data are reserved, which enables more accurate and reliable analysis.

Secondly, with the result of effective distance matrix, we project the countries in the two-dimensional space with MDS method. It helps to explore the relationship between countries intuitively. The world pattern of population migration is uniform and decentralized, and conversely, it shows more imbalance in trade network, where more countries are clustering.

Thirdly, we rank the countries with minimum reverse flow skill. The result indicates that the international trade networks do not have a strong hierarchy. However, countries or areas have a strict hierarchy for global migration.

To be more specific, this paper uses real data to calculate the actual relationship. Obtaining true relationships based on real data is an important concept in the study of complex systems, especially needing a comprehensive perspective [15–18]. The relationship in between the countries is extremely important, and this article is just a preliminary attempt to analyze the scenario.

Acknowledgements. We appreciate comments and helpful suggestions from Prof. Yiming Ding, Jiang Zhang, Hongbo Cai, and Dr. An Zeng. Thank Ms. Liqian Lang for some help in data processing. This work was supported by Chinese National Natural Science Foundation (71701018) and (61673070), National Social Sciences Fund, China (14BSH024), and Beijing Normal University Cross-Discipline Project.

References

1. Soytas, U., Sari, R.: Energy consumption and GDP: causality relationship in G-7 countries and emerging markets. Energy Econ. **25**(1), 33–37 (2003)
2. Nieh, C.C., Lee, C.F.: Dynamic relationship between stock prices and exchange rates for G-7 countries. Q. Rev. Econ. Finan. **41**(4), 477–490 (2001)
3. Meyer, D.F., Masehla, T.M., Kot, S.: The relationship between economic growth and economic development: a regional assessment in South Africa. J. Adv. Res. Law Econ. **8**(4(26)), 1377–1385 (2017)
4. Samargandi, N., Fidrmuc, J., Ghosh, S.: Is the relationship between financial development and economic growth monotonic? Evidence from a sample of middle-income countries. World Dev. **68**, 66–81 (2015)
5. Baylis, J., Smith, S., Owens, P. (eds.): The Globalization of World Politics: An Introduction to International Relations. Oxford University Press, Oxford (2017)
6. Hidalgo, C.A., Klinger, B., Barabási, A.L., Hausmann, R.: The product space conditions the development of nations. Science **317**(5837), 482–487 (2007)
7. García-Pérez, G., Boguñá, M., Allard, A., Serrano, M.Á.: The hidden hyperbolic geometry of international trade: World Trade Atlas 18702013. Sci. Rep. **6**, 33441 (2016)
8. Noulas, A., Scellato, S., Lambiotte, R., Pontil, M., Mascolo, C.: A tale of many cities: universal patterns in human urban mobility. PLoS ONE **7**(5), e37027 (2012)
9. Krings, G., Calabrese, F., Ratti, C., Blondel, V.D.: Urban gravity: a model for inter-city telecommunication flows. J. Stat. Mech: Theory Exp. **2009**(07), L07003 (2009)
10. Kahouli, B., Maktouf, S.: Trade creation and diversion effects in the Mediterranean area: econometric analysis by gravity model. J. Int. Trade Econ. Dev. **24**(1), 76–104 (2015)
11. Ramos, R., Surinach, J.: A gravity model of migration between the ENC and the EU. Tijdschrift voor economische en sociale geografie **108**(1), 21–35 (2017)
12. Park, J.: Diagrammatic perturbation methods in networks and sports ranking combinatorics. J. Stat. Mech. Theory Exp. **2010**(04), P04006 (2010)

13. Clauset, A., Arbesman, S., Larremore, D.B.: Systematic inequality and hierarchy in faculty hiring networks. Sci. Adv. **1**(1), e1400005 (2015)
14. Borg, I., Groenen, P.J.: Modern Multidimensional Scaling: Theory and Applications. Springer Science & Business Media, New York (2005)
15. Brockmann, D., Helbing, D.: The hidden geometry of complex, network-driven contagion phenomena. Science **342**(6164), 1337–1342 (2013)
16. Chen, Q., Huang, Z.G., Wang, Y., Lai, Y.C.: Multiagent model and mean field theory of complex auction dynamics. New J. Phys. **17**(9), 093003 (2015)
17. Yan, X.Y., Zhao, C., Fan, Y., Di, Z., Wang, W.X.: Universal predictability of mobility patterns in cities. J. R. Soc. Interface **11**(100), 20140834 (2014)
18. Simini, F., González, M.C., Maritan, A., Barabási, A.L.: A universal model for mobility and migration patterns. Nature **484**(7392), 96 (2012)

Special Operations Forces: A Global Immune System?

Joseph Norman(✉) and Yaneer Bar-Yam(✉)

New England Complex Systems Institute,
210 Broadway, Cambridge, MA 02139, USA
joe.w.norman@gmail.com, yaneer@necsi.edu
http://www.necsi.edu

Abstract. The use of special operations forces (SOF) in war fighting and peace keeping efforts has increased dramatically in recent decades. A scientific understanding of the reason for this increase would provide guidance as to the contexts in which SOF can be used to their best effect, and when conventional forces are better suited. Ashby's law of requisite variety provides a scientific framework for understanding and analyzing a system's ability to survive and prosper in the face of environmental challenges. We have developed a generalization of this law to extend the analysis to systems that must respond to disturbances at multiple scales. This analysis identifies a necessary tradeoff between scale and complexity in a multiscale control system. As with Ashby's law, the framework applies to the characterization of successful biological and social systems in the context of complex environmental challenges. Here we apply this multiscale framework to provide a control theoretic understanding of the historical and increasing need for SOF, as well as conventional military forces. We propose that the essential role distinction is in the separation between high complexity fine scale challenges as opposed to large scale challenges. This leads to a correspondence between the role SOF can best serve and that of the immune system in complex organisms—namely, the ability to respond to fine-grained, high-complexity disruptors and preserve tissue health. Much like a multicellular organism, human civilization is composed of a set of distinct and heterogeneous *social tissues*, each with its own distinct characteristics and functional relationships with other tissues. Responding to disruption and restoring health in a system with highly diverse local social conditions requires an ability to distinguish healthy tissue from disruptors and to neutralize disruptive forces with minimal collateral damage, an essentially complex task. Damage to social tissue, either through the growth of malignant forces or large-scale intervention by conventional forces, leads to cascading crises that spread beyond the initial location of disruption. To prevent such crises, the healthy functioning of social systems must be maintained by responding to disruptive forces while they remain small. SOF have the potential to mitigate against harm without disrupting normal social tissue behavior. Three conditions for SOF to fulfill such a role are identified: (1) distinctive capabilities of special operators that enable unmediated interaction with local cultures and peoples, (2) persistent presence and

A. J. Morales et al. (Eds.): ICCS 2018, SPCOM, pp. 486–498, 2018.
https://doi.org/10.1007/978-3-319-96661-8_50

embeddedness to foster cultural attunement and mutual trust, and (3) local autonomy and decision-making of SOF to achieve requisite variety for sensing and acting on fine-grained disturbances. We point out the inapplicability of traditional hierarchical control structures for high-complexity local tasks, which require a decentralized control architecture. This analysis suggests how SOF might be leveraged to support global stability and mitigate against cascading crises.

1 Executive Summary

Special Operations Forces (SOF) provide war fighting capabilities that complement conventional forces. A conceptual framework is needed to clarify and differentiate the role of SOF within the larger military system to aid decision-makers in identifying when it is necessary and appropriate to utilize SOF and when conventional forces are better suited.

Here, we propose a correspondence between the role SOF may serve and that of the immune system in complex organisms.

In organisms, the immune system is composed of many semi-autonomous components and is responsible for sensing and acting on fine-grained, high-complexity disturbances that may harm the growth and functioning of healthy tissue. It must differentiate between *self* and *other* by having an intimate knowledge of the character of local tissue, detecting agents and behavior that pose a threat. When functioning effectively the immune system eliminates harmful agents without disrupting the normal behavior of healthy tissue.

Much like an organism, global civilization is composed of a collection of diverse *social tissues*, each with its own distinct form, way of living, and functional role within larger communities. SOF are uniquely positioned to develop the knowledge and capabilities to distinguish healthy social tissue and detect and mitigate threatening forces.

Conventional forces, by contrast, are well suited to external threats and their use in societal challenges may damage social fabric, leading to disrupted and vulnerable states.

Three conditions must be met to enable SOF to eliminate threats while preserving the health of social tissue:

1. Distinctive capabilities and advanced training of special operators – Advanced cultural and language competencies and experience in making difficult decisions in the face of uncertainty enable unmediated interaction with local people. When necessary, they can strike with exacting force.
2. Persistent presence and enduring engagements – Repeated and habitual interaction with local communities at both the individual and institutional levels provides the opportunity to develop necessary cultural attunement.
3. Local autonomy and decision-making – Acting on nuanced information relevant to local conditions engenders the ability to stem local threats. Locally embedded SOF must have the freedom to behave semi-autonomously, making many decisions independently of SOF located elsewhere or central command structures. This requires avoiding the tendency to become bureaucratized.

The correspondence between the potential role of SOF and the immune system in organisms is formalized via multiscale control systems theory and a *complexity profile* analysis. Ashby's *law of requisite variety* sets the lower bound of complexity a system must possess to survive and prosper. SOF can provide essential complexity at fine scales, as does the immune system in biological systems.

This correspondence suggests that SOF are uniquely equipped to serve as a global immune system, acting before threats rise to the level of crises, and preserving healthy and diverse social tissue functioning. Future policy decisions will determine the degree to which these unique SOF capabilities are developed and leveraged.

2 Introduction

Throughout history warfare has involved both large-scale conventional conflict, in which armed combatants seek to gain physical advantage over their adversaries, as well as less conventional operations which focus on high-value targets or seek to achieve a desired effect through indirect means. The latter have come to be known as *special operations*, and they are typically carried out by small groups and individuals with distinctive skills, creativity, and often equipment. In 1987, the United States Special Operation Command (USSOCOM) was established to oversee the nation's special operations forces (SOF) for both independent and joint operations.

SOF have their origin in military practice, and only recently have attempts been made to articulate a theory of the role of SOF [1–3] as a subset of a larger *theory of warfare*. These efforts highlight the need for a framework that provides guidance to decision-makers about when and how to utilize SOF to their greatest effect, and when other options are more appropriate.

Here, we present a theory of SOF motivated by mathematical and physical necessity and grounded in complex systems science. We propose a correspondence between the functional role of SOF and that of the immune system in complex biological organisms, and a parallel correspondence between conventional forces and the neuromuscular system. According to this theory, SOF play a vital role in sensing and acting in fine-grained, high-complexity environments, complementing conventional forces that sense and act at larger scales.

The theory brings military theory into contact with a body of scientific knowledge and inquiry about the behavior of complex systems and, crucially, the conditions under which they are able to survive and prosper.

The remainder of the article is divided into four sections. First, relevant concepts in the theory of multiscale control systems are reviewed and summarized. Second, these concepts are applied to clarify the functional complementarity of the immune and neuromuscular systems in complex organisms. Third, the functional role of SOF is couched in this theory and brought into correspondence with that of the immune system. Finally, strategy, policy implications, and implementation challenges are discussed.

3 Multiscale Control Systems

3.1 The Law of Requisite Variety and Its Limitations

In 1956, W. Ross Ashby formalized in the study of control systems what is known as *the law of requisite variety* [4]. In short, the law of requisite variety sets the minimum number of behaviors, or 'variety', a system must have to survive and prosper in a given environment. As the number of distinct situations a system encounters increases, the variety of its behavioral repertoire must also increase in order to achieve desired outcomes—or as Ashby put it: "variety destroys variety". This concept is illustrated in Fig. 1. If a system has little variety or is overly-constrained while being exposed to a large variety of stressors (i.e. a complex environment), it will sooner or later fail to achieve desired outcomes. In this article, we will use the terms *variety* and *complexity* interchangeably. Thus an environment with *high-complexity* is one with a large variety.

The theory of control systems traditionally deals with systems at a well-defined *scale* of relevant behavior, and abstracts away details that are presumed not to be of concern due to the nature of the system or the method of control. For example, if one wanted to construct a robot that could catch a baseball, one need not be concerned with the atomic vibrations ongoing within the baseball, but rather its relevant macroscopic properties like mass, location, and trajectory, and corresponding control variables like joint angles and positions.

In contrast, living systems are exposed to environments with stressors and complexity on multiple relevant scales that must be effectively managed to achieve self-regulation and good overall system health. For example, as organisms we are exposed not only to traffic as we cross the street, but also to microscopic organisms that may find our bodies to be suitable homes within which to replicate themselves to our detriment. These two sources of stress exist at scales separated by several orders of magnitude, and our bodies therefore have different strategies in controlling for their potentially harmful effects.

Thus, the law of requisite variety per se is not enough to account for how, say, an organism achieves self-regulation in a complex environment with multiple scales of impinging forces and stressors. Both *variety* and *scale* must be considered for good control in complex multiscale environments [5,6].

3.2 Scale/Complexity Tradeoff

There is an inherent tradeoff between the scale and complexity of behavior in any system. In order for large-scale behaviors to occur, a large number of components must work coherently or in coordination. Consider, for example, the muscle tension that ultimately gives rise to the movement of a limb. If only one or a small number of muscle fibers become engaged, the scale of the force will be small, and the limb will express essentially no behavior. However if many muscle fibers become engaged at once, a larger scale force is produced, and the limb will change its position—a large-scale behavior is induced through the coherent activity of many parts.

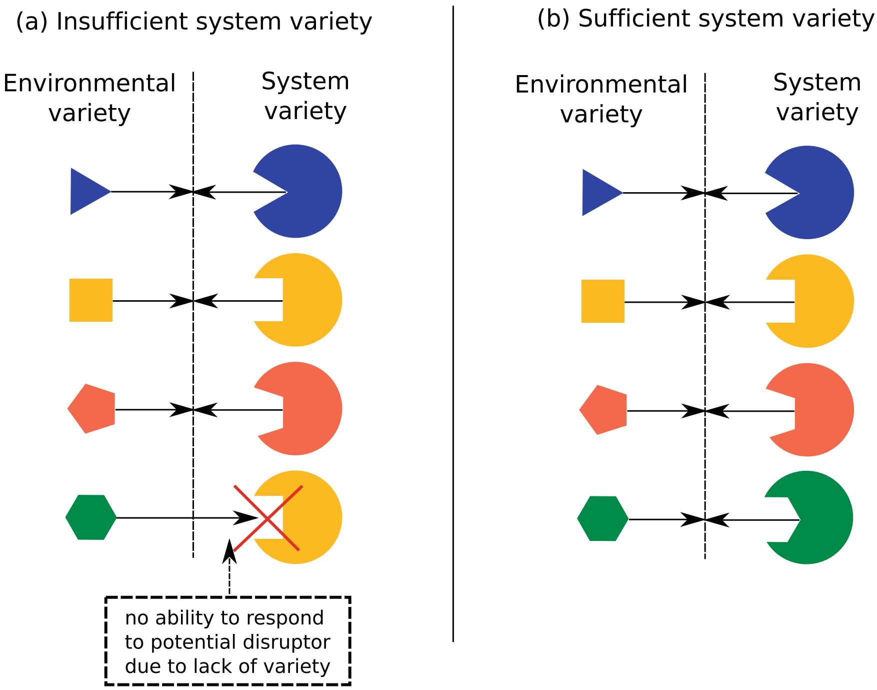


Fig. 1. Requisite variety. Panel (a) shows a system (right) being exposed to environmental disturbances (left). The variety of the environment is greater than the variety of the system, as there are 4 unique disturbances but only 3 unique responses. The system lacks the ability to respond to the green hexagon, which will disrupt it. Panel (b) shows a similar case, but where the system variety matches the variety of potential disturbances—the system has the requisite variety to respond to all potential disturbances. Systems with variety greater than that of their environment also possess requisite variety.

The flip-side of achieving large-scale effects through coherent behavior of many components is that those components are not free to behave independently, but are constrained by the role they play in the large-scale behavior. This decreases the variety, or complexity, that can be expressed by the system at small scales.

We can quantify a system's variety at a given scale. In a system with N components that behave independently, the number of states the system can achieve is the product of the number of states each component can take. For instance if each component can take on 2 states (a *binary* system), the number of total possible states is 2^N. More generally, if n_i denotes the number of states component i can take on, the total number of states of the system is $\prod_i n_i$. The components must act in concert in order to achieve large-scale effect. This necessarily reduces their degree of independence, and the number of states of the system, or variety, is less than $\prod_i n_i$. In other words, due to constraints that prevent each component from behaving independently, the actual variety is less than its maximum would be without constraint. Constraints are indicative of underlying structures that enable large-scale behaviors and variety.

We can summarize the essential tradeoff as follows: *demand for variety at large-scales necessitates the reduction in variety at smaller scales.*

For a given system, this tradeoff can be captured and summarized via the *complexity profile* (Fig. 2) which represents variety, or complexity, as a function of scale [6–10]. The 'shape' that the complexity profile of a system takes on reflects its structure and behavior and identifies the scales over which they are present. When the smallest components of a system behave essentially independently, there is a maximal amount of variety at the fine-scale. However, as we move to larger scales, the independent behavior of all these parts 'average out' and we observe no large-scale behavior. When all of the smallest components move together coherently across the entire system, like the atoms in a baseball when thrown, we find behavior at larger scales, with variety varying minimally across scales, and variety at fine-scales being reduced dramatically compared to the case of component independence. If you know the flight path of one atom in the baseball, you know them all. For objects like complex organisms, we observe a mixture of these two modes. The variety at the smallest scales remains quite

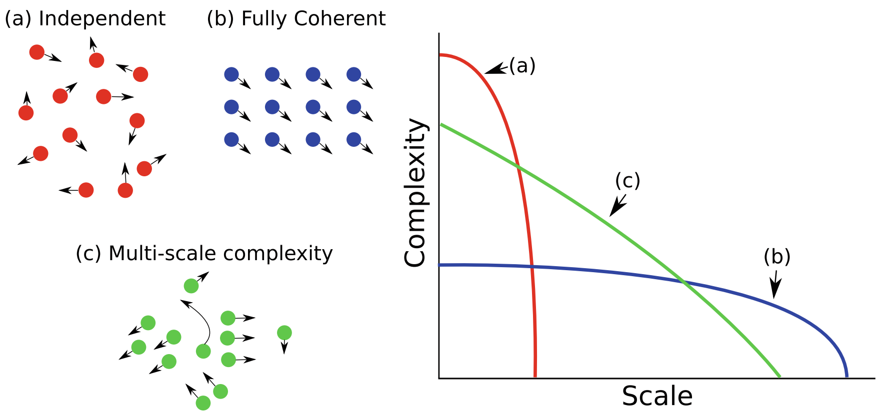


Fig. 2. Complexity profile. The complexity profile maps system complexity, or variety, as a function of scale. Three example cases are shown. In the case of system components behaving with complete independence (a), there is high-complexity at the finest scales, but variety quickly drops off to zero, and no behavior is observed at larger scales. A system in which the components are fully coherent (b) has substantially reduced variety at small scales. Intuitively, this is because if we know the behavior of one component, we know them all (i.e. they are constrained). However, this enables large-scale behavior, as the components behave in concert. Systems with multiscale complexity (c) have both fine-scale complexity (though less than case (a)) and can produce large scale behaviors (though not as large-scale as case (b)). This is achieved by some components behaving in a coherent and coordinated fashion, while others are free to behave semi-autonomously.

high (though, less than in the case of complete independence), while many of the components are coordinated into larger structures that reduce their independence, but achieve larger scale behaviors.

The different multiscale behaviors also require different control structures to enable actions to be performed in response to environmental challenges or conditions [5]. For large scale behavior, hierarchical control is appropriate. This is because, on the sensing side, large, coherent external events are detected, and irrelevant details are filtered out for high-level decision making. On the action side, unified decisions can be projected to a large number of agents who behave in concert to achieve a large-scale effect. In contrast, for behavior that responds to fine-scaled, high-complexity challenges that don't require large-scale response, hierarchical control is inappropriate and insufficient. The details lost as information ascends the hierarchy are precisely the ones relevant to small scale decisions. Moreover, the projection of operational directives from high-levels is necessarily insensitive to these low-level details, constraining agents' behavior and preventing them from responding and adapting to the local context. Instead, distributed networks of agents with minimal hierarchical constraints leave small groups and individuals able to make decisions semi-autonomously, retaining sensitivity to local information and enabling adaptation and response to these locally-relevant variables.

4 The Immune and Neuromuscular Systems

A familiar example of a multiscale control system with clear differentiation of scale and function is the physiological system that combines the neuromuscular and immune systems in multicellular organisms, such as ourselves. The key differences between the form and function of these two systems lend insight into the nature and role of SOF, and that of conventional forces.

The neuromuscular and immune systems operate concurrently in order to achieve overall system health by responding to disturbances at different scales. The neuromuscular system detects large-scale events and structures in the external environment—dodgeable cars, walkable paths, climbable trees, fall-offable cliffs—and generates coordinated behavior of the gross physical structure of the body in order to leverage opportunities or mitigate harm. These faculties operate in the 'Newtonian' macroscopic environment of everyday life to avoid physical damage and provide the resources necessary for physiological function.

The immune system serves a different, but equally important, function. It is distributed and embedded throughout the body and its tissues. It contains a variety of cell types that behave with a large degree of independence—behaviors are not constrained to achieve large-scale coherence as in the neuromuscular system. One of its essential roles is to differentiate *self* and *other* at the cellular and sub-cellular scales in order to promote the flourishing of healthy tissue and eliminate or neutralize threats when detected. The cells of the immune system sense and act locally, without direct instruction from centralized command structures, though 'training' and other functions are centralized in lymph nodes and bone marrow.

The distinction between self and other is not genetic, but rather associated with healthy functioning. Consider a cancerous cell and a bacterium that aids in healthy digestion. The former would be appropriately identified as *other*, despite sharing its genome with the host, and the latter *self*, because of its functional harmony with the host.

The body is organized into a collection of heterogeneous tissues and organs which serve various functions that complement one another, forming a self-consistent whole. A well-functioning immune system promotes a healthy system by minimizing the potential for disruption of local tissue. This is a critical point: the integrity of the functional tissue is preserved via the action of a healthy immune system, and a disruption to *any* of the tissues in the body disrupts their role and can lead to cascading effects throughout the whole system, including those cells not directly affected by the disruptor, and perhaps organism death.

Notably, the immune system does not *direct* the tissue or instruct its behavior explicitly, but rather creates the conditions in which it can express its distinct form without harming itself or other tissue.

The essential differences between the neuromuscular and immune systems are the scales over which they operate, and the degree of independence of components that determine the scale and variety. The neuromuscular system operates with a great degree of coordination among its parts, limiting variety at fine-scales and producing it at large-scales. The components of the immune system, by contrast, behave more independently, making decisions locally and maintaining variety at the fine-scales, enabling sensing and acting that preserves good tissue functioning and avoids disruption.

The immune system cannot catch a baseball, and the neuromuscular system cannot eliminate a bacterium. The *only* way the neuromuscular system could effectively combat a micro-disruptor would be through the destruction of functional tissue, an action with irreversible and often system-wide consequences.

5 Multiscale Military Theory and the Functional Role of SOF

Much like an organism, our global civilization is composed of a set of distinct *social tissues*, each with unique character, mode of internal operation, and interfaces with other tissues. Healthy, well-functioning social tissues have internal behaviors that sustain the individuals composing them, such as agriculture, goods production, trading and markets, health services, social gatherings and celebrations, as well as fruitful external interactions with other social systems such as the buying and selling of commodities, products, and services.

When healthy and functional social tissue is disrupted, opportunities are created for malignant forces to gain footholds and grow. This dynamic can be seen, for instance, in the unintended consequences of the invasion of Iraq, which created the opportunity for terrorist networks and other harmful actors to increase their power and influence as normal life was disrupted and power vacuums were created. Moreover, the harm and risks generated by the growth of malignant forces are not confined to the local area where they first manifest.

Because of our global interconnectedness and interdependence, effects cascade causing disruption in other tissues, leading to a domino effect with no straightforward mechanism to halt the expanding impacts [11,12]. The recent and ongoing migrant crisis in Europe and beyond provides an example of one form such cascading effects can take.

Global interdependency means any large-scale military intervention, by virtue of disrupting the normal functioning of society, will generate both local and non-local unintended consequences even when desired effects are achieved. This is not to suggest that large-scale action is never necessary, but the potential for generating new crises must be weighed carefully whenever it is considered as an option. In many cases, action that does not disrupt local, healthy social behavior is possible, but it requires the right action and agent.

The parallels of the effects of tissue disruption in organisms and in sociocultural systems highlights the need for a 'sociocultural immune system'—a fine-grained system for sensing and acting on environmental disturbances at scales smaller than conventional forces are able. In this regard, conventional forces can be likened to the large-scale neuromuscular system in organisms. Acting instead at a small-scale presents the possibility of maintaining healthy social tissue and allowing it to flourish. Just as for the immune system, this is not a matter of differentiating 'native' and 'foreign', but understanding whether an agent is disruptive to overall health.

SOF are uniquely positioned to fulfill this role, possessing the requisite personnel, skills, and training. For this to be realized, policies that impact SOF must be such that they enable their unique capabilities in meeting the high-complexity demand of local cultural systems. We identify three conditions that must be satisfied in order for SOF to serve such a role: special operators with advanced training and distinctive capabilities, persistent presence and enduring engagements, and local autonomy and decision-making. We discuss each in turn.

5.1 Distinctive Capabilities

Much like the cells in the immune system have special forms and functions to fulfill their roles, the distinctive capabilities of special operators enable them to operate in highly complex sociocultural environments. Advanced language and cultural training allows unmediated interaction with local peoples. Special operators' experience in making decisions in the face of uncertainty allow them to operate in ill-defined 'gray zone' conditions.

The need to produce special operators with distinctive capabilities highlights the role of SOF's high-selectiveness, and emphasizes the necessity of advanced training in language and culture in addition to combat. These values are articulated in the SOF truths "Humans are more important than hardware" and "SOF cannot be mass produced" [13]. Preparing special operators to interact directly and make difficult decisions in complex psychosocial, sociocultural, and kinetic environments must be a priority of SOF and their enabling agencies.

5.2 Persistent Presence

The immune system is embedded throughout the tissues of the body to develop and maintain sensitivity to the character of local tissue and respond rapidly to disruptors [14]. Similarly, persistent presence of SOF allows for nuanced relationships to unfold over time, and for cultural attunement to be developed at both the individual and institutional levels. SOF embeddedness engenders an understanding of normal conditions and a sensitivity to changes in those conditions and whether they pose a threat. Moreover, presence is necessary for applying rapid and effective action to achieve desired effects with minimal disruption. Just as SOF must recognize 'self' in multiple contexts, local cultures must not react to SOF as a foreign entity, i.e. mutual trust must be present, developed through shared history.

Policies should enhance continuity of interaction between SOF and a given sociocultural system even, or especially, when there is no immediate or visible threat. The only way to prevent the growth of malignancies is to be present and active *before* they grow. This is reflected in the SOF truth "competent SOF cannot be created after emergencies occur" [13] and Admiral William McRaven's oft-cited comment that one "can't surge trust" [15].

5.3 Local Autonomy and Decision-Making

As the cells of the immune system sense, decide, and act locally in a decentralized manner, being fine-tuned to the character of their local tissues, so too must SOF have the ability to sense, decide, and act locally using their nuanced understanding and experience.

The semi-autonomy of SOF is necessary for requisite variety to be achieved in interfacing with high-complexity, fine-grained environments and disruptors.

In human systems, these disruptors manifest at the psychosocial and sociocultural scales. It is possible to take effective action at these scales to eliminate harmful agents without disrupting healthy social tissue functioning. This becomes impossible as the scale of a malignancy grows larger: social tissue will inevitably be damaged by both the malignancy itself and any large-scale force applied in response.

When the decision-making agent is both far removed from and insensitive to the the local context, as well as receiving multiple information streams about which decisions must be made, the sensitivity, nuance, and understanding of local SOF is lost. Consequently, the ability to stem malignant forces while they remain small in scale is diminished, and the likelihood of disrupting a social system either accidentally or out of necessity as the scale of harmful actors grows larger increases.

To enable SOF to act without disrupting social tissue, the institutions overseeing SOF must not over-constrain their behavior. As policy- and decision-makers look increasingly to SOF to overcome complex challenges, it is critical that they do not become overly-bureaucratized.

Imperatives that are communicated to SOF must be guided by their role as a protector of local tissue function. Protections from the potential for harm to local tissues, i.e. by civilian collateral damage from operations, must be instituted in a way that retains local autonomy. Detailed instructions on how to carry out missions will prevent them from behaving as necessary for success in high-complexity environments. In technical terms, placing too many constraints on their behavior will reduce their variety below the (requisite) threshold for sensing and acting on fine-grained disruptors. The consequences of this are twofold: (1) SOF will lack the ability to sense and eliminate threats while they remain small, and (2) disruption and destruction of healthy social tissue becomes inevitable as malignant forces grow and large-scale intervention becomes the only means of engagement.

6 Challenges and Implementation

While SOF are uniquely positioned to fulfill an immune-system-like function, there remain significant challenges to successful implementation. Here we summarize some of these challenges.

Developing SOF who are both culturally and linguistically competent, as well as able to execute reconnaissance and surveillance and direct action missions demands significant investment in training and preparation. Moreover, for any individual operator there is a tradeoff in developing proficiency in any given domain. However, SOF must be able to perform the entire range of activities, from sensing nuanced changes in social conditions to taking actions to eliminate harmful disruptors, to preserving social tissue health.

Fundamental limitations on individual capabilities lead to a need for diversification of roles of SOF. This is manifest already in different types of SOF, as it is in the immune system which uses cells of various types, each of which serves particular roles that complement one another. The relative levels of activity for the different cell-types vary depending on circumstance; some cells primarily sense tissue conditions and detect disruptors, while others act to confine and eliminate harmful agents once they are identified. During an infection rapid clonal reproduction (replication) of effective types occurs. Similarly, SOF may embrace and develop specialization of expertise, and should be flexible enough to adapt force size and composition in response to changing circumstances.

Maintaining the mental health of special operators must be a priority, and appropriate support systems should be put in place for this. The high complexity of tasks translates into the psychological symptoms of stress, depression and burnout, common in a high complexity society more generally but surely for SOF. Moreover, adapting to diverse local contexts creates challenges when switching to home and family environments, a potential component of post traumatic stress disorder (PTSD). This is a challenge for both the SOF and their families.

Giving special operators a significant degree of autonomy presents challenges and risks that are distinct from those of conventional command-and-control systems. Care must be taken to ensure social tissue is not damaged by unintentional friendly fire, collateral damage or intentional ‘rogue operators’. The potential for disfunction is not unlike auto-immune disorders in complex organisms and the immune system has developed mechanisms for prevention, though no mechanism is failure proof. Local feedback systems including multiple specialized roles rather than centralized control ones must be in place that put checks on the actions of operators.The structure of these feedback mechanisms must be the subject of intensive study.

Rapid growth in recent years has led to institutionalization of SOF using concepts that may be incorrectly adopted from command control military traditions. Bureaucratization runs counter to the ability of SOF for performing the functions we have identified. Rather then enabling SOF function as it grows, institutionalization may result in undermining the effectiveness of SOF as it becomes more like conventional forces. Alternative structures must be developed. They may be inferred from fundamental complex systems analyses, including correspondence with immune system functions or well designed experimentation.

Institutional structures and relationships between SOCOM and other enabling agencies, including those within DoD, and other departments of the executive branch such as the State Department, need to be carefully considered. For example, how the agenda of an ambassador of a given region and local SOF should interrelate is an open question. If command and control structures do not appropriately interface with SOF, their unique capabilities will not be utilized effectively. This includes knowing when and, crucially, when *not* to utilize SOF to achieve a desired effect. This article is intended to contribute to this clarification.

7 SOF in the 21st Century

There is no doubt that as a global civilization we will continue to face fine-grained, high-complexity disruptors that have the potential to grow into larger-scale malignancies. The only way to combat this is to promote and enable the flourishing of healthy social tissues. Multiscale control systems theory makes clear the need for an immune-like system embedded within human social systems. It must be sensitive to and embedded within high-complexity psychosocial and sociocultural environments to make decisions locally based on understanding of a given social system, its nuances, and distinctive qualities.

Like the various tissues arranged into functional organs throughout the body, cultures and social systems do not all look, behave, or function alike. Part of a global strategy for the 21st century must be the recognition that cultures can not simply be ‘exported’ or ‘projected’ onto others without pushback, and that behavioral diversity at the collective scale is a natural and healthy part of our human civilization. SOF possess the unique organizational capabilities to be sensitive to the healthy behavior of these diverse ‘social tissues’, while providing the direct and indirect action capabilities to neutralize malignant forces when identified.

Moving forward, a major part of the SOF repertoire must include relationship building. Interpersonal relationships with local individuals form the basis of understanding necessary to discern between harmful and beneficial (or neutral) forces to social health. The ability to perceive and understand local tensions, grievances, typical and atypical interactions, customs, and other nuanced features can serve to generate solutions before the normal functioning of healthy social tissue is threatened. The highly-complex and fine-grained nature of this endeavor makes it an unsuitable role for conventional forces – they can not sense nor act on such a fine scale. A focus on direct action is important when specific disruptors have been identified, and not otherwise.

SOF is uniquely positioned to serve as a global immune system, keeping the diverse set of social tissues healthy, and reserving large-scale intervention for when it is necessary.

Acknowledgement. We thank Charles Flournoy and Philip Kapusta for helpful comments and discussion.

References

1. Kiras, J.D.: A theory of special operations: these ideas are dangerous. Spec. Oper. J. **1**, 75–88 (2014)
2. Yarger, H.R.: 21st century SOF: Toward an american theory of special operations. Technical report, DTIC Document (2013)
3. Spulak Jr., R.G.: A theory of special operations: The origin, qualities, and use of SOF. Technical report, DTIC Document (2007)
4. Ashby, W.R.: An Introduction to Cybernetics. Chapman & Hail Ltd., London (1956)
5. Bar-Yam, Y.: Multiscale variety in complex systems. Complexity **9**, 37–45 (2004)
6. Allen, B., Stacey, B.C., Bar-Yam, Y.: An information-theoretic formalism for multiscale structure in complex systems. arXiv preprint arXiv:1409.4708 (2014)
7. Bar-Yam, Y., Harmon, D., Bar-Yam, Y.: Computationally tractable pairwise complexity profile. Complexity **18**, 20–27 (2013)
8. Bar-Yam, Y., Cares, J.R., Dickmann, J.Q., Glenney IV, W.G.: Multiscale representation phase I. Final Report to Chief of Naval Operations Strategic Studies Group (2001)
9. Bar-Yam, Y.: Making Things Work. Knowledge Press, Cambridge (2004)
10. Bar-Yam, Y.: Dynamics of Complex Systems. Addison-Wesley Reading, Boston (1997)
11. Bar-Yam, Y.: Complexity rising: from human beings to human civilization, a complexity profile. In: Encyclopedia of Life Support Systems (EOLSS), UNESCO. EOLSS Publishers, Oxford (2002)
12. Lagi, M., Bar-Yam, Y., Bertrand, K.Z., Bar-Yam, Y.: Accurate market price formation model with both supply-demand and trend-following for global food prices providing policy recommendations. Proc. Natl. Acad. Sci. **112**, E6119–E6128 (2015)
13. USSOCOM: SOF truths. Accessed 8 Feb 2016
14. Matzinger, P.: The danger model: a renewed sense of self. Science **296**, 301–305 (2002)
15. Trulio, D.: You can't surge trust—Insights from the opening of the aspen security forum (2012). Accessed 27 July 2012

Engineered Complex Adaptive Systems of Systems: A Military Application

Bonnie Johnson(✉)

Naval Postgraduate School, Sandy Spring, MD 20860, USA
bwjohnson@nps.edu

Abstract. Tactical warfare is complex [1]. It requires agile, adaptive, forward-thinking, fast-thinking, and effective decision-making. Advancing threat technology, the tempo of warfare, and the uniqueness of each battlespace situation, coupled with increases in information that is often incomplete and sometimes egregious; are all factors that cause human decision-makers to become overwhelmed [2]. Automated battle management aids become part of a solution to address the tactical problem space—to simplify complexity, increase knowledge, and formulate quantitative analyses of decision options. The other part of the solution is engineering an adaptive architecture of distributed weapons and sensors that can act independently or as collaborative systems of systems. This paper proposes a systems approach to the complex tactical problem space. The approach is based on a complex systems engineering strategy that views the decision space holistically in the context of capability enablers for managing future distributed warfare assets as complex adaptive systems of systems.

Keywords: Complex adaptive systems of systems · Battle management aids
Decision-making

1 Introduction

The ability to optimally manage distributed warfare assets for collaborative operation significantly increases our military advantage. Recent studies have pointed to an increasing speed of warfare, emerging threat capability and numbers, and data overload from a growing number of sensors and networks. This results in challenges to human decision making when faced with a complex decision space, multitudes of information, and the fast reaction time required. Automated battle management aids (BMAs) have the potential to reduce timelines, increase decision confidence, and optimize warfare resources. This paper describes a systems engineering approach to conceptualizing and designing BMAs for future military warfare missions. A systems approach views BMAs holistically in the context of capability enablers for managing future distributed warfare assets as complex adaptive systems of systems (CASoS).

Automated BMAs are computer-aided decision support systems that are meant to enhance and improve tactical decisions. BMAs may improve decisions by: speeding up the decision process; providing greater confidence in the knowledge that decisions are based on; developing more decision options; providing greater understanding of decision consequences; developing options with greater probability of success; and

A. J. Morales et al. (Eds.): ICCS 2018, SPCOM, pp. 499–506, 2018.
https://doi.org/10.1007/978-3-319-96661-8_51

improving the optimization of resource usage. The military currently uses BMAs to share and process data to develop operational pictures. However, this paper is focused on conceptualizing BMAs as envisioned for the future for military operations. A systems approach integrates the analytic and synthetic methods, encompassing both holism and reductionism [3]. It emphasizes the interdependencies and interactions between elements within a system and between systems and their external environments [4]. This paper proposes the necessity of following a systems approach for the conceptualization and engineering of future automated BMAs.

2 Decision Aids for Battle Management

As preparation for conceptualizing automated BMAs, this section characterizes the types of decisions made for battle management. It discusses how BMAs may be used to support human decision-makers within a tactical environment. It introduces the concept of "decision complexity" and the role of BMAs to address tactical complexity.

2.1 Battle Management Decisions

Tactical operations involve a great variety of battle management decisions. Most decisions involve the use or placement of warfare assets which include platforms (ships, aircraft, submarines, etc.), weapons, sensors, communication devices, and people [5]. There are four domains of warfare decisions: temporal, spatial, proactive/reactive, and the domain of rules and policies. Each of these domains affects the decision-making process and can lead to increased decision complexity.

Planned or proactive decisions include positioning forces (ships, battlegroups, aircraft, etc.), stealth operations, offensive attacks, and denying enemy operations through jamming or other force measures. Examples of reactive or responsive decisions include defending against an active threat, moving platforms into a defensive posture, retreating from a threat environment, and assessing battle damage. Effective battle management must recognize when proactive or reactive decisions need automated support.

The nature of military decisions shifts over time and can be viewed as hierarchical. Strategic decisions have a longer time horizon and consider high level objectives—sometimes spanning years. Planning-level decisions have a shorter time horizon and are proactive even when arranging a defense. Tactical decisions, the main focus of battle management, have the shortest time horizon and involve very near-term planning or proactive decisions as well as reactive decisions in response to enemy actions. Consistency is desired among the three temporal decision domains to effect compatibility among tactical, planning, and strategic decisions. Likewise, plans and strategies need to support effective tactical warfare and reflect major changes in tactical threat environments. Automated BMAs should be designed to support a hierarchical decision paradigm as well as one that supports and adapts to varying decision time horizons.

One of the results of the hierarchical temporal decision domain is a set of rules and policies that guide tactical decisions. These rules are one of the methods by which nearreal-time decisions can align with longer term plans and strategies. The rules and policies support effective tactical decisions that are consistent with the higher objectives. BMAs could support dynamic decision-making across the temporal and hierarchical domains to enable consistency among levels; consideration of how changes at various levels might affect other levels; and effective promulgation of guidance across levels.

A fourth way to categorize battle management decisions is by spatial domain; such as space, air, sea, underwater, and land. Threats vary greatly in each of these environments. Likewise, warfare systems are developed to address specific threat types which naturally reflect their spatial environment. Naval battlegroups must address threats in all spatial domains, and at times, simultaneously. BMAs have the potential to address this complexity through gains in cross-spatial-domain situational awareness and through the development of decision alternatives that prioritize missions and engagement strategies.

Ultimately, the battle management decision space fluctuates from simple to complex as operations range from peace-time to multi-domain threat encounters. Examples of changes to the problem space that affect the complexity of the decision space include: battle tempo (or reaction time), the number of simultaneously-occurring threats (or battle events), the severity of the consequences of battle events, the heterogeneity of threats (due to threat type or spatial domain), and the scope of the event or events (in terms of area or population affected). All of these operational factors translate into multi-dimensional variables that comprise a "decision space." As the decision space complexity increases, military human decision-makers may become overwhelmed. At this point, having automated BMAs in place, can support effective decision-making.

2.2 Automated Aids to Support Human Decisions

The amount of information in the battlespace has increased due to more sensors, networks, participants, reach-back and intelligence. Human decision-makers become overwhelmed with information and shortened decision times. Automated BMAs are a necessary capability required for effective tactical decision-making.

Automated decision aids, or "machines," can support human decision-makers in a number of ways. Three models for human-machine decision-making interaction are include manual, semi-automated, and fully automated [5]. The manual decision-making model encompasses situations in which humans collect and "store" relevant information as well as perform the decision analysis (processing and decision-making), in their heads. This model implies a fairly simple and straightforward decision space in which the amount of data and number of variants is manageable manually. In the semi-automated model, the human decision-maker can rely on machines to manage, store, fuse, and process the input information to display decision analytics. Decision analytics may consist of knowledge of the battlespace and threats, course-of-action (COA) options, and quantitative measures of expected event successes and consequences. Finally, in the fully automated model, the role of the human is to monitor the automated machine decision processes and to override or change decisions when necessary.

It is important to establish the appropriate mechanism for the type of decision being made. In general, decision-making can be performed manually when the problem space is relatively simple and the number of factors to be considered and the amount of information is manageable by the human decision-maker. For some types of decisions, a semi-automated HMI mechanism is most appropriate. This is effective for more complex decision spaces with potentially critical or dire consequences; requiring the support of automated BMAs, but with significant human involvement. A fully automated HMI is appropriate for decision spaces that are complex in terms of large amounts of information that must be processed and fused; but very straightforward in terms of the types of decisions being made. Fully automated decision modes are for peace-time operations where decisions do not have dire consequences or for highly complex operations where the decision reaction time is too compressed for humans. Fully automated decision modes are appropriate when there is very high confidence in the knowledge of the situation. For example, when it is known with high confidence that a tracked object is in fact an enemy threat target.

A future goal for BMAs is to have the capability to select the appropriate decision model for the given decision space. Perhaps a flexible decision-making architecture can accommodate all three human-machine models and apply them as needed. The superstructure, itself, would be monitoring the decision space and evaluating what kinds of decisions needed to be made and then determining the appropriate interaction between the human and machine to make each decision.

2.3 Battle Management: A Complex Endeavor

Battle management operations are complex [6]. The tactical environment can range from peaceful to highly dangerous with a multitude of varied threats from many different directions. This translates into a complex decision space for battle management. The "state" of the decision space must flexibly shift from linear and straightforward during non-threat operations, to highly nonlinear and multi-varied during combat.

Characteristics of a complex problem space include: complex objectives, complex environments and/or operations; adaptation; collective behavior; and unpredictable outcomes of decisions. Each of these characteristics are inherent to tactical operations [6]. The battlespace presents multiple objectives that are generally inconsistent and changing. Military systems must weigh their individual battle objectives, such as self-defense, against force-level missions which may include area defense, stealth operations, or defense of specific assets. Complex operations are required as adverse and widely varying environments result in changing target priorities and multiple cross-spatial domain missions. Adaptation is a required characteristic of warfare systems as they respond to the complex and changing threat environment. Military operations must adapt effectively to threats to improve their chances of survival and meet tactical and strategic goals. The collective behavior of distributed warfare assets must be properly orchestrated to avoid collisions and friendly-fire incidents; and ideally benefit from their cumulative contributions. Finally, the unpredictable outcomes of tactical decisions ranging from misfires to misidentifications to misassessments of battle damage, result in a problem space made more complex through inaccurate knowledge and a ripple effect of actions and unforeseen consequences.

Automated BMAs have the potential to support human decision-makers by characterizing the level of complexity in the operational environment and translating this knowledge to the decision space. Ideally, a complete and accurate "picture" of the battle space will provide situational awareness to the decision space. BMAs could monitor the picture and develop assessments of the complexity characteristics of the problem space. This knowledge could support effective and timely use of decision aids as well as enable the effective interplay of human and machine decision-making.

3 A Systems Approach

A high complexity task requires a system that is sufficiently complex to perform it [1, 7]. Tactical operations present highly complex environments that translate into complex tasks for warfare assets. This section explores a complex systems approach to implementing automated BMAs into military operations to effectively address tactical problem spaces. The previous section characterized the battle management problem space in terms of decision-making; made the distinction between decisions made by humans and how automated decision aids can support those decisions; and characterized battle management complexity. This section introduces a way of thinking about the problem space as a means to conceptualize and ultimately implement a systems solution.

3.1 A. Viewing Warfare Assets as Resource Systems

The first step in a systems approach is to "view" the problem and solution spaces in terms of systems. For tactical warfare, this begins with viewing warfare assets as resource systems. Defining assets (such as: ships, aircraft, submarines, weapon systems, sensors, communication devices/networks, and jammers) as systems, allows them to be considered as resources and viewed in terms of their functions, performance, behavior, structure, and interfaces. It enables quantitative analyses to be performed based on their characteristics such as location, status, and expected capabilities. As operations grow in complexity, automated BMAs could perform analyses to determine the effective use of warfare resources when multiple objectives exist that overlap and conflict. Warfare resource utilization could, with the aid of BMAs, include forming collaborations among systems to enable system of systems capabilities to better address complex missions.

"Multidimensionality is probably one of the most potent principles of systems thinking. It is the ability to see complementary relations in opposing tendencies and to create feasible wholes with infeasible parts [4]." By viewing the battlespace as a set of interacting systems, the ability to exploit their multidimensionality supports collaborative force-level behavior that spans spatial and temporal domains. It enables layered defense and integrated fire control strategies involving distributed weapons and sensors. BMAs can provide the quantitative analysis to determine collaborative resource utilization when complex multidimensional objectives exist.

3.2 Viewing Battle Management Holistically

Complex tactical environments require a holistic perspective to manage warfare resources from a force level. As the environment becomes more complex, events are occurring more rapidly and in parallel. The numbers of decisions are increasing as are the number of courses of actions required. More demands are being made on the finite set of warfare resources and their missions, objectives, and courses of action are becoming more interrelated. Gaining a "holistic" understanding of multiple threats and missions as well as the possible options for addressing them as well as the possible consequences provides a more effective military response and may be required to effectively address demanding threats. The idea of battlespace perspective can be characterized as "decision scope," or setting a boundary around the problem space and solution space. A more holistic decision scope includes an area or theater and all threats and warfare resources in this geospatial area. A narrower decision scope may only include a particular threat and a particular platform and its associated assets.

Establishing decision scope is both a limiting factor and a necessary enabler. Tactical decisions become more interdependent and "messy" in terms of cause and effect as the operational environment becomes more complex [8]. Making a particular weapons engagement decision or sensor tasking decision is simpler when there is one threat to kill or one area of interest to view. However, narrowing the decision scope to firing a single weapon system or managing the sensors on one ship, loses its overall force-level effectiveness when several tactical missions need to be addressed or many threats need to be prioritized and engaged. The principle of "holism," applied to decision-making in this context involves including "simultaneously and interdependently as many parts and levels of the system as possible [8]." In other words, widening the scope of the decision space to perhaps consider a tactical area or theater. Determining the decision scope is a decision in itself. The goal is to design future force architectures that support a flexible decision scope that can widen as force-level missions become more complex and might benefit from distributed warfare asset collaboration.

Once a tactical military force faces a complex operational problem space, future automated BMAs could establish a more holistic and wider decision scope and support resource management at both the platform and force levels. Ultimately a variety of automated BMAs could support resource usage at different levels. BMAs supporting specific sensors and weapons could be orchestrated by a higher level BMA architecture. Thus a system of BMA systems could be implemented.

3.3 Viewing the Decision Space as a System

The decision space can be viewed as a system. By taking a systems approach, it enables the definition of a boundary, inputs and outputs, functionality, performance, and structure. Knowledge (or situational awareness) of the battlespace is developed and maintained as the problem space (or operational tactical picture). It includes tracked threat objects as well as terrain, weather, defended assets, and all other physical entities in the real world. A resource picture must also be developed that includes up-to-date status, health, readiness, and projected capabilities of the warfare assets. The problem space and resource pictures comprise the primary inputs to the decision space.

The boundary of the conceptual decision space system surrounds the decision architecture, and the decision analytics which include decision aids, assessments, prioritizations, alternatives generation, and overall decision management. The primary function of the decision space system is to develop decision alternatives. These alternatives provide recommendations to manage the warfare resource assets. Examples include sensor tasking, COAs, weapon scheduling, and the movement of platforms. Secondary functions include estimating the confidence levels associated with decision alternatives and the many types of analyses that feed into alternatives. Examples include prioritizing threats, wargaming possible consequences, estimating sensor error, estimating knowledge accuracy and completeness, evaluating operational complexity, and recommending optimum human-machine decision-making interaction. Outputs of the conceptual decision space system could include decision alternatives, estimations of predicted consequences, estimated probabilities of success and failure; and confidence levels associated with source information, options, and knowledge in general.

3.4 Solution Space: Complex Adaptive Systems of Systems

A final step in this overview of a systems approach to BMAs, is the conceptualization of the solution space. With a goal of enabling a tactical response to a complex threat space, the solution space consists of the effective use of distributed warfare assets/resources. The solution must change in time and adapt as the threat environment changes. At times an offensive action is the best option, at other times a single platform can address the threat, and yet at other times, a multitude of offensive, defensive, collaborative, and autonomous actions may be required, both parallel and in series. The ability of the solution space to shift seamlessly from simple to complex operations, thus changing the nature of its system state, is a challenging requirement.

This paper conceptualizes the solution space as a CASoS in which the distributed warfare resources interact as systems of systems, exhibiting emergent (force-level) behavior, and adapting to the changing operational environment. This class of systems is a required solution to effectively address complex tactical problem spaces. Engineering future warfare systems to behave as CASoS requires a decision architecture and solution space of automated BMAs that provide the following three capabilities [9]:

1. Adaptive relationships: an adaptive intelligent architecture enables agile interrelationships among the constituent systems that comprise an ultimately adaptive SoS that can respond to a changing complex environment.
2. A system of intelligent constituent systems: the adaptive emergent behavior of the CASoS is governed by the self-management of the distributed constituent systems to collaborate or act independently as the complex situation dictates.
3. Knowledge discovery and predictive analytics: key to the engineered CASoS is the ability to gain and maintain shared situational knowledge of the environment and the distributed constituent systems. The knowledge is analyzed to prioritize missions; develop tasks and courses of action (adaptive responses to the problem space); and to develop “what-if” and “if-then” predictive scenarios to shape the synthesis of future intelligent decision and adaptive SoS relationships.

The decision space must support the conceptualized CASoS solution space. The decision space for this complex application can be thought of as a system of BMA systems with holistic force-level management decision aids supporting the orchestration of lower-level BMAs concerned with specific resource or platform systems. The holistic-level BMAs could manage the problem space information and focus on high-level concerns such as evaluating the level of complexity, establishing decision scopes, and recommending human-machine decision interactions. All of this requires BMAs, an adaptive architecture, warfare resources that are taskable, and a command and control culture that supports this systems approach.

4 Conclusions

The battle management problem space is complex and will only continue to grow in complexity with the addition of more sensors, information, unmanned threats, non-state adversaries, and advances in technology. To stay ahead of this problem space, a complex solution must be conceptualized and eventually realized to facilitate fast-acting and highly responsive warfare utilization. A systems approach provides a method for addressing the multidimensional and adaptive decisions required by offering holism, a systems perspective and the definition of the decision space as a system of systems. It frames the problem as a CASoS and highlights the need for a decision architecture that enables adaptive relationships, intelligence at the system level, shared knowledge, and predictive analytics. The effective use of automated BMAs in support of human decision-making provides the foundation for the CASoS solution space.

References

1. Bar-Yam, Y.: Making Things Work: Solving Complex Problems in a Complex World. NESCI Knowledge Press (2004)
2. Zhao, Y., Kendall, A., Young, B., Baumgartner, W.: Big Data Architecture and Analytics for Improving Combat Identification. Naval Postgraduate School Report (2015)
3. Checkland, P.: Systems Thinking, Systems Practice. Wiley, Hoboken (1993)
4. Gharajedaghi, J.: Systems Thinking: Managing Chaos and Complexity. Morgan Kaufman, Burlington (2011)
5. Johnson, B., Green, J., Canfield, W.: Gaining naval battlespace through automation. In: Proceedings of the National Fire Control Symposium (2001)
6. Young, B.: Complex systems engineering applications for future battle management and command and control. In: Proceedings of the 18th International Command and Control Research and Technology Symposium (2012)
7. Braha, D., Minai, A., Bar-Yam, Y.: Complex Engineered Systems, Springer (2006)
8. Jackson, M., Keys, P.: Towards a system of systems methodologies. J. Oper. Res. Soc. **35**(6), 473–486 (1984)
9. Johnson, B.: Towards a theory of engineered complex adaptive systems of systems. In: Proceedings of IEEE Systems Conference (2018)