Characterizing Student Theory Building in Computational Modeling

Hillary Swanson, hillary.swanson@usu.edu, Utah State University
Bruce Sherin, bsherin@northwestern.edu, Uri Wilensky, uri@northwestern.edu, Northwestern University

Abstract: It is widely agreed that engaging students in authentic science practices is important for science education. Theory building is a central practice of science. Today, many scientists build theory through computational modeling. This paper characterizes the nature of student theory building in the context of computational modeling activities. Using a fine-grained grounded analysis, we identified theory-building moves in one student's goal-directed modeling in a block-based microworld. We present moves enacted during phases of model building, testing, debugging, and sense-making using a segment of transcript from the student's theory-building activity, which was focused on modeling a flu epidemic.

Introduction

It is widely agreed that engaging students in authentic science practices is important for science education (Duschl, Schweingruber, & Shouse, 2007). Theory building is a central practice of science. Many scientists build theory by constructing computational models that, when run, produce outcomes that can be explored and compared with experimental findings (Foster, 2006). A number of research programs have explored ways of engaging students in theory building through computational modeling. diSessa (1995) describes a case where high school students re-invented F=ma through their development of computational models. Wilensky and colleagues have investigated student construction of models of complex systems phenomena such as predator-prey dynamics, using the NetLogo computational modeling environment (Wilensky, 1999; Wilensky & Reisman, 2006). Recent work in this tradition has examined student construction of models using NetTango Web (Horn, Baker & Wilensky, 2020), a block-based interface to NetLogo. The present work builds on this tradition by examining the nature of student theory building in the context of computational modeling activities. It seeks to characterize elements of theory building supported by block-based microworlds.

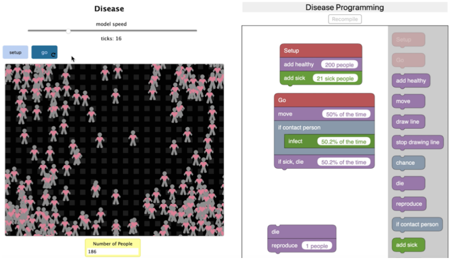
Theoretical Foundations

We define scientific theory building as a family of practices through which scientists and students systematically refine theoretical knowledge artifacts, including laws, models, explanations, constructs, and categories (Swanson, 2019). As these artifacts are refined, thinking is refined. Our perspective aligns with constructivist frameworks that view the construction of new knowledge as a refinement of prior knowledge (diSessa, 1993). It also aligns with constructionism (Papert, 1980), which argues that learning happens best through the construction of public artifacts, such as computational models. In our work, we seek to characterize students' theory building by describing the moves through which they refine their computational models. These are moves through which their ideas are systematized and formalized, and brought into alignment with evidence and observation. We are focused on identifying these moves for two reasons. First, we view them as productive resources that can be leveraged and developed by instruction toward expertise in theory building. Second, we are interested in mapping a gradient from less to more expert theory building, and in so doing, locating student theory building on a continuum with the theory building of scientists. In this paper, we focus specifically on characterizing the *process* of student theory building, leaving the science *learning* that results to other papers.

Methods

We present results from an analysis of data taken from a larger study focused on supporting middle school student engagement in different approaches to scientific theory building, including the construction of computational agent-based models. To make computational agent-based modeling more accessible, we are designing block-based microworlds using the NetTango interface to NetLogo. NetTango makes the computational power of NetLogo accessible to authors by using a block-based programming language curated to a particular phenomenon. NetTango blocks are not a full programming language, but domain-specific blocks relevant to the modeled phenomenon. Previously called *semantic blocks (*Wilkerson-Jerde & Wilensky, 2010) and now called *domain blocks* (Wagh et al., 2017) the blocks are, from a student's point of view, primitive elements of code that represent agents' actions, which can be combined to model a specific phenomenon. We are designing domain-block libraries for simulating complex systems phenomena and studying how children use the blocks to engage in scientific theory building. In this study, we ask the question *"What is the character of student theory building in the context of computational modeling?"*

To address this question, we tested NetTango modeling microworlds with middle school students (ages 12-14) during one-on-one 1.5-hour task-based interviews. During each interview, the student had full command of a laptop featuring an agent-based microworld. The interviewer guided them through tasks and questions from a semi-structured protocol, which introduced the features of the microworld and then prompted the student to model a particular phenomenon. This study focuses on an interview with Sage, a 13-year-old who had just started 8th grade at a public school in her small Midwestern city. Sage explored the *Spread of Disease* model, shown in Figure 1, below. Figure 1 shows the microworld with a model of Ebola built by Sage. The black box to the left is the *world* which depicts the activity of the agents that are programmed to behave according to the rules specified by the model. The *setup* and *go* buttons are controlled by procedures (red blocks) that the user must drag from the block library (far right) into the modeling field (middle) and then define by connecting with blocks (purple, grey, and green), such as *move*, *if contact person*, and *infect*.



```
Setup
add healthy 200 people
add sick 21 sick people

Go
move 50% of the time
if contact person
  infect 50.2% of the time
if sick, die 50.2% of the time

die
reproduce 1 people
```

Figure 1. The *Spread of Disease* microworld featuring Sage's model of Ebola.

Sage's interview was recorded using video, audio, and screencast technology. The audio recording was transcribed. A fine-grained grounded analysis was applied to both the screencast and transcript to identify theory-building moves that Sage enacted (Glaser and Strauss, 2017). First, the screencast of Sage's interview was reviewed and times were noted during which she engaged in building models for particular diseases, namely Ebola, the flu, and a zombie apocalypse. These episodes were then marked on the transcript, which was read for evidence of theory-building moves, which were named and organized into phases of model building. These were: 1) building, 2) testing, 3) debugging, and 4) sense-making. For the purposes of illustrating the moves she enacted, the moves were used to characterize the episode of Sage's modeling in which she tried to model a flu epidemic. The episode was temporally decomposed into steps of model building, testing, debugging, and sense-making. The episode is presented in the Findings section as a piecewise trajectory. The transcript associated with each step is followed by a description of the smaller theory-building moves in which Sage engaged, highlighted in italics.

Findings

The grounded analysis of the three episodes revealed 46 theory-building moves across the four phases. We present a narrative account of a segment of Sage's modeling activity to illustrate how she engaged in building, testing, debugging, and making sense of her model. We offer a temporal decomposition of the first three minutes of the 10-minute trajectory through Sage's flu modeling activity, highlighting her theory-building moves in italics.

Modeling the Spread of Influenza

Sage is seated at a desk in an office, the interviewer sits at her left. She is looking at a laptop screen on which the *Disease Spread* modeling microworld is open. She has been exploring the microworld for the last 40 minutes. Her exploration began by trying out combinations of blocks and watching the resulting activity in the *world*. She then built and refined a model of Ebola (Figure 2), at the interviewer's request. Following this, the interviewer asks Sage how she would modify the code in the Ebola model, to model the flu.

Step 1: Building the Initial Flu Model

Sage looks at the model of Ebola and considers how she might modify the code to represent the flu.

> Sage: If it was flu, hmm, it's less deadly [...] Like, I don't know, like 10%.

To match her understanding of the flu, Sage *purposefully selects a parameter value*, decreasing the probability that a sick agent will die, from 50.2% to 10%, and then to 5.5%. She appears to be *drawing on prior knowledge* she has about the difference in deadliness of the flu *as compared* with Ebola. She *approximates the parameter value* and shows an *awareness of the limits of her knowledge*, qualifying her choice with the words “I don’t know.”

Step 2. Testing the Initial Flu Model

The interviewer asks Sage how she thinks her modification is going to influence the model outcome.

> Sage: I think the epidemic is going to spread like much further because, um, people like when people are moving around and infecting people, they're like, they're, they're not, they don't have like a 50% chance of dying every time. [...] And so people weren't, people were dying faster than they were coming in contact with people instead of-

Sage *makes a prediction for the outcome of the model run*, and *explains her prediction* by *comparing her expectation with results of previous runs* of her model of Ebola. She *explains the aggregate-level outcomes of those runs as the result of agent-level interactions*. Her reasoning is that for Ebola, “people were dying faster than they were coming in contact with people,” and the disease disappeared from the world before it could turn into an epidemic. For the flu, she reasons that the epidemic will spread because it is less deadly and therefore sick people will be in the world long enough to infect healthy people. She appears to be *drawing on knowledge* she constructed while modeling Ebola, that if the disease is deadly to infected individuals it will kill them off before they can spread it and decimate the population. Sage presses “go,” running the model to *test the effect of her new parameter* and *observing* as the model runs. She watches as the number of people in the world decreases as healthy people become infected and sick people die. This outcome stands in contrast with her Ebola model, where all of the initially sick people died in the first few ticks.

Step 3: Making Sense of the Initial Flu Model

The interviewer asks Sage if she thinks this model looks different from her model of Ebola.

> Sage: Yeah. Um, I think the percentage of population is like going down as you can see.
>
> Interviewer: Ah, so what did you change in the um, the flu case?
>
> Sage: Well, people died less and it spread less. And so, we went from like 221 people, 41 people because it's, um, it's very infectious, but it's like, like people don't die a lot. But when there are a lot of people then people start dying. [...] And so just people keep dying and then eventually [...] So, because no one's recovering, you can't, I didn't have it set to recover.

Sage tests the model and *describes its outcome*: that the population is decreasing over time. She *compares the results of this run with results of the previous runs* of the Ebola model, where fewer people died, because the sick people died before the disease could spread. She turns to the flu model-run results, *referencing numerical data* that she *calculated* by subtracting the value given in the box titled “number of people” from the total initial population (the population has decreased by 41). She *explains the aggregate-level outcome of the flu model in terms of agent-level behavior*. She explains that the flu is very infectious but less deadly to those infected. If there are a lot of people, some percentage of the population will die, so the population decreases. Through her sense-making, Sage *arrives at a powerful conclusion about complex-systems dynamics*: when a disease is less deadly to infected individuals, infected people die less and the disease spreads to more of the population. If there are enough people in the population, many people will ultimately die. Sage acknowledges that the population may have decreased too much in her model, because the only option for sick people is to eventually die, as she left out the “recover” block. Sage *connects the agent behavior with the model code* and engages in debugging, identifying this as a *problem with the model*.

Step 4: Debugging the Initial Flu Model

The interviewer asks Sage if she thinks as many people die in a flu epidemic as shown in her model.

> Sage: Probably not. It's probably because I have the infectivity high and people aren't recovering. But let's say you recover like, I don't know, 50%? No, like 75 but that's going to be way too high. Yeah, 75 and less infectious, and this will be hard to start your... Wait, no, that's recover, wait. Yeah. Infect like 10%.

At the interviewer's prompting, Sage *assesses the reasonableness of her model's results*. She explains the unexpected *outcome at the aggregate level as the result of both an incorrect parameter setting* and *incorrect agent behavior*. She engages in debugging, *modifying the code to address the problem* by adding the "recover" block to the model and setting the probability of recovery to 73.1%. She also debugs the model by *modifying a parameter to address the problem*, reducing the probability of infection from 50% to 10.5%. Following this, Sage continues to work on her model for another 7 minutes, working through an additional 18 steps of testing, making sense of, and debugging the model. All the while her goal was to find a way to get the epidemic to sustain itself and work more like her image of the flu epidemic. Over the 10 minutes of the activity, Sage demonstrated many moves that may be productive foundations for developing expertise in areas of theory building, including model building, testing, debugging and sense-making.

Discussion

This study characterizes student theory building in the context of computational agent-based modeling. Working in the microworld with the support of the interviewer, Sage engaged in theory building by constructing, testing, debugging, and making sense of a computational model. By identifying and characterizing Sage's theory-building moves, we identified resources she brings to her learning that may serve as foundations for developing expertise in theory building. In this way, we make a contribution to literature concerned with identifying the resources novices bring to their learning that can be foundations for the development of scientific expertise. The work makes a contribution to the larger project of characterizing the nature of student engagement in different forms of scientific theory building by characterizing student engagement in computational agent-based modeling. By identifying the smaller moves of which this approach to theory building is composed, our findings are foundational for the development of science curricula and assessments that promote and capture rich student engagement in scientific theory building.

References

diSessa, A. A. (1993). Toward an epistemology of physics. *Cognition and instruction*, *10*(2-3), 105-225.

diSessa, A. A. (1995). Designing Newton's laws: Patterns of social and representational feedback in a learning task. In R.-J. Beun, M. Baker, & M. Reiner (Eds.), *Dialogue and Interaction: Modeling Interaction in Intelligent Tutoring Systems.* Berlin: Springer-Verlag, 105-122.

Duschl, R. A., Schweingruber, H. A., & Shouse, A. W. (Eds.). (2007). *Taking science to school: Learning and teaching science in grades K-8* (Vol. 500). Washington, DC: National Academies Press.

Foster, I. (2006) 2020 computing: a two-way street to science's future. Nature 440(7083):419

Glaser, B. G., & Strauss, A. L. (2017). *Discovery of grounded theory: Strategies for qualitative research*. Routledge.

Horn, M.S., Baker, J. & Wilensky, U. (2020). NetTango Web [Computer Software]. Evanston, IL: Center for Connected Learning and Computer Based Modeling, Northwestern University. https://netlogoweb.org/nettango-builder

Papert, S. (1980). Mindstorms: Computers, children, and powerful ideas. *NY: Basic Books*.

Swanson, H. (2019, April). *Cultivating a theoretical turn-of-mind.* Paper presented at the annual meeting of the American Educational Research Association, Toronto, Canada.

Wagh, A., Cook-Whitt, K., & Wilensky, U. (2017). Bridging inquiry-based science and constructionism: Exploring the alignment between students tinkering with code of computational models and goals of inquiry. *Journal of Research in Science Teaching*, *54*(5), 615-641.

Wilensky, U. (1999). NetLogo. http://ccl.northwestern.edu/netlogo/. Center for Connected Learning and Computer-Based Modeling, Northwestern University, Evanston, IL.

Wilensky, U., & Reisman, K. (2006). Thinking like a wolf, a sheep, or a firefly: Learning biology through constructing and testing computational theories—an embodied modeling approach. *Cognition and instruction*, *24*(2), 171-209.

Wilkerson-Jerde, M., & Wilensky, U. (2010). *Restructuring Change, Interpreting Changes: The DeltaTick Modeling and Analysis Toolkit*. Paper presented at the Constructionism 2010 Conference, Paris.

Acknowledgements

This work was supported by the National Science Foundation (1842375).