COMPUTATIONAL MODELING IN THE ELEMENTARY SCIENCE CLASSROOM

By

Amanda Catherine Dickes

Dissertation

Submitted to the Faculty of the

Graduate School of Vanderbilt University

in partial fulfillment of the requirements

for the degree of

DOCTOR OF PHILOSOPHY

in

Learning, Teaching & Diversity

May, 2017

Nashville, Tennessee

Approved:

Pratim Sengupta, Ph.D.

Richard Lehrer, Ph.D.

Leona Schauble, Ph.D.

Gautam Biswas, Ph.D.

To the dedicated teachers and students whose commitment, curiosity and enthusiasm made this work possible.

ACKNOWLEDGEMENTS

This research was funded in part by the National Science Foundation. I would like to express my sincerest gratitude to my advisor, Pratim Sengupta, who years ago saw a potential in me that I had yet to see in myself. He has pushed my thinking in innumerable ways, contributed sage advice on design and writing and motivated me to pursue questions beyond the edge of knowledge. I will be forever grateful for his guidance, support and mentorship. I would also like to extend my heartfelt thanks to my other committee members – Rich Lehrer, Leona Schauble & Gautam Biswas – and a special thanks to Doug Clark. Each has challenged, encouraged and refined my thinking and provided a model for the scholar and teacher I would like to become.

During my tenure as a doctoral student I have been privileged to learn alongside some of the greatest thinkers – and people – in our field. I would like to especially thank my fellow "Pratim-mates", Amy and Kara, who contributed countless hours in support of my work. Their friendship and encouragement sustained me throughout this process. I would also like to thank my game-night gang – Robin, Sean, Mayumi, Adrian and Mario – in whose company I have erected wonders of the ancient world, built civilizations and destroyed the One Ring. Between rolls of the dice, they provided me with needed perspective and constantly challenged my conceptions of teaching and learning.

Finally, I would like to thank my family – Mom, Dad, Rachel, Tres, Chris, Nathan and Maddy – for their unconditional love and support, enthusiastic encouragement and shelter from the occasional storm. They have been my biggest cheerleaders and this accomplishment is as much theirs as it is mine.

TABLE OF CONTENTS

APPENDIX

LIST OF TABLES

Chapter 4

LIST OF FIGURES

Chapter 4

CHAPTER 1

INTRODUCTION

In the decades since their emergence in the late nineteen seventies, personal computers have transformed nearly every aspect of human society, culture and economy (diSessa, 2001; Maloy & LaRoche, 2014). A "once in several centuries innovation" (Simon, 1983), computers have revolutionized disciplines such as science, engineering and communication (diSessa, 2001; Woolf, 2010; Vee, 2013) and, more recently, have continued to diversify into such fields as art (Peppler & Kafai, 2005), achitecture (Vee, 2013) and history (Maloy & LaRoche, 2014). Notable scholars have attributed the impact of computing and computers to what Papert (1980) has called their richness of material. Computers, as Papert notes, are "object[s]-to-think-with", providing users with powerful and concrete ways to reason through novel problems.

Given this impact on culture and society, it is unsurprising that leading educational scholars have long argued for *computational thinking*, an analytic problem solving and design approach fundamental to computing (Wing, 2006), to be an essential focus of K12 curriculum (diSessa, 2001; Papert, 1980; Repenning, Basawapatna & Klymkowsky, 2013; Sengupta, Kinnebrew, Basu, Biswas & Clark, 2013; Wilensky, 1995; Wing, 2008). Computational thinking has now been incorporated as an essential concept for science, technology, engineering and math (STEM) education in the Next Generation Science Standards (NGSS, 2013). However, studies have shown that curricular integration of computational thinking and modeling is a complex and challenging endeavor (diSessa, 1991; Sherin, diSessa & Hammer, 1993; Guzdial, 1994) which involves the introduction and adoption of new literacies (e.g., programming) to both teachers and students, alongside disciplinary ideas and practices that students already find challenging to understand (Sengupta, Kinnebrew, Basu, Biswas & Clark, 2013).

This dissertation presents a set of three studies that attempts to investigate different dimensions of the problem of *merging* computational modeling and thinking with K12 classrooms in the elementary grades. While "computing" can mean a broad range of activities and practices, in this dissertation I focus on two specific forms: modeling and programming (or coding). In this section, I first present an overarching theoretical framework grounded in diSessa's (2001) notion of *computational literacy*, and explain the three intertwined dimensions – social, material and cognitive – along and across which computational literacy must be understood. I will then introduce the design and goals of the studies, and explain the dimensions addressed in each study. Taken together, these studies illuminate several less-understood aspects of learning, teaching and educational design of computational thinking within elementary science curriculum.

Computational Literacy

diSessa (2001) has argued that computational technologies only become truly revolutionary upon their transformation from a *material intelligence* into a *literacy*. Material intelligence indicates the deployment of skills and capabilities with computational technologies. Literacies, however, involve more than tool use. They allow people to "negotiate their world" through their impact on a wide variety contexts, both mundane and profound (diSessa, 2001). diSessa argued that while individuals greatly benefit from material intelligences, it is through literacies that knowledge is both influenced and generated.

Material intelligences transform into literacies when they become infrastructural to a society's communicative practices (diSessa, 2001). That is, there is widespread ability to compose and interpret with that technology, what diSessa terms “two-way” literacy. Social forces play a large role in the development of literacy and at the time of writing, diSessa was uncertain how long it would take for societies to consider computational literacy powerful enough and valuable enough to be worth the considerable effort of teaching it to everyone. Today, it can be argued that the communicative technologies of computation – code – are now infrastructural to modern society. Programming structures much of contemporary communications, including email, word processing and social networking. However, the ability to *read* and *write* in code is still not widespread. This trend is changing. Educational researchers, business leaders and the federal government now recognize computational literacy as a “basic skill” necessary for economic opportunity and social mobility (Computer Science for All, 2016) and are focused on teaching coding in K12 as computer science (Computer Science for All, 2016) as well as teaching coding as part of K12 STEM curriculum (Sengupta, Dickes, Farris, Karan, Martin & Wright, 2015; Wilensky, Brady & Horn, 2014).

Three pillars of literacy. diSessa (2001) and other scholars (Vee, 2013) have argued that literacy of any form, and therefore, computational literacy, involves an interplay between *material*, *social* and *cognitive* dimensions. While diSessa referred to them as “pillars”, I refer to them as dimensions, because of their deeply intertwined nature. Investigating each of these dimensions as well as their interactions in K12 settings can help us understand the nature of computational literacy and how to support it in K12 classrooms. The material dimension of literacy involves creating and modifying symbolic systems (e.g., coding) as well as physical computing (e.g., physical computing using Arduino and 3D printers). Developing expertise along this dimension in turn involves developing expertise along the cognitive dimension (e.g., learning to use programming commands and computational abstractions such as data structures).

At the same time, the social dimension is omnipresent; these manipulations and transformations of materials occur in specific social contexts (the social dimension of literacy) where complex social forces of innovation, adoption and interdependence transform material intelligences into a literacy (diSessa, 2001; Street, 1984; Vee, 2013).

Computing certainly involves creating symbolic representations within a generalized programming environment; but it is also an interpretative act that draws on knowledge and practices acquired in specific social contexts (diSessa, 2001; Sengupta, Kinnebrew, Basu, Biswas & Clark, 2013). Even computational abstractions, which Wing (2006) referred to as the key of computational thinking, usually become evident in the form of contextualized expressions (Sengupta, Kinnebrew, Basu, Biswas & Clark, 2013). Context is not merely a container for computing; it also shapes the form and practice of computing (Sengupta & Shanahan, *In Press*). For example, domain specific modeling languages (DSML), in which programming languages not only represent the core computing abstractions but also the disciplinary contexts in which it is to be used, are now becoming increasingly popular in software engineering, in contrast to generalized programming languages that only use generalized computing abstractions (Schmidt, 2006). Similarly, the ubiquity of embedded and distributed computing (e.g., the Internet of Things, Arduino microcontrollers, 3D printing) have enabled seamless integration of the virtual and the physical worlds (Sengupta, Dickes, Farris, Karan, Martin and Wright, 2015; Blikstein, 2015).

What does this mean for research on integrating computational thinking and modeling with science curricula in the elementary grades? At the broadest level, this is the question my dissertation seeks to answer. In the classroom, in light of diSessa's three pillars (dimensions), answering this question involves examining the complex interplay and negotiations between computational and non-computational representations (material), reasoning and discourse (cognitive and social), and the emergent classroom micro-culture (social). My goal is to study the complex interplay and negotiations within and across these dimensions that are involved in the development of computational literacy in elementary science classrooms.

Overview of this work

Individually, each chapter of this work investigates how computational modeling, specifically computational modeling using agent-based computational models, can be integrated with elementary science curriculum to support the co-development of scientific and computational literacy. Collectively, they explore how students and the classroom teacher make use of forms of activity that integrate computational modeling with other forms of scientific modeling (physical, embodied and mathematical) in an attempt to understand how *computation* and *computational modeling* can become the "language" of practice in the elementary science and math classroom.

Chapter 2. The work reported in this chapter examines the close-interplay between the material and cognitive dimensions of literacy by investigating the forms of reasoning fourth graders utilized to develop more expert-like explanations of predator-prey relationships and population change due to natural selection after interacting with an agent-based model.

In the study, ten fourth graders representing the top five and bottom five performers in the class interacted with an agent-based computational model designed specifically for the study. The primary researcher interviewed each student during their interaction with the model. Interview questions and activity scaffolds were designed to uncover any conceptual entities activated during manipulation of the model interface, and how those conceptual entities developed during the course of interaction. Analysis of student interviews revealed that conceptual entities termed *registrations*, or initial conceptions, were activated during students' early explorations of the model and, upon continued interaction with the technology, were productively used to generate more sophisticated explanations of population-level behavior, such as population increase due to camouflage, by connecting agent-level attributes, behaviors and/or interactions to corresponding emergent behavior.

Chapter 3. The work presented in Chapter 3 further elaborates on the interplay between the material and cognitive dimensions of literacy and investigates the forms of reasoning and knowledge that developed in a third-grade classroom during interaction with a computational model of predator-prey interactions. Additionally, this chapter extends the work conducted in Chapter 2 and investigates how computational modeling is enhanced through its integration with other material forms, specifically with scientific modeling, including embodied modeling and the development of mathematical representations of change over time. Classroom discourse and the emergent classroom micro-culture were an additional component of this work. The role of the teacher in re-shaping the structure of activity, and how those re-shapings influenced the knowledge that developed during activity are discussed.

The study design was focused around three phases of activity 1) Participation in an embodied modeling activity of butterflies foraging for nectar 2) The generation of mathematical inscriptions of energy change over time and finally 3) Interaction with two different computational models of predator-prey dynamics in an ecology of butterflies, flowers and birds. Classroom instruction was shared between the lead researcher and the classroom teacher, however, the classroom teacher was instrumental in guiding and constraining learning objectives in different activities.

This work provided a possible pathway for integrating computational modeling to support the development of scientific and computational literacy in the elementary classroom. This work highlights two important characteristics of such integration: the importance of integrating computational modeling with other forms of activity, namely embodied modeling, to support the co-development of computational thinking and scientific knowledge in the classroom. And, the role the classroom micro-culture plays in the how computation is taken-up as an emerging literacy in the development of scientific practice.

Chapter 4. Chapter 4 further develops the work conducted in Chapter 2 and 3 and focuses on how computational modeling and programming are integrated into existing elementary science curriculum over a longer period of time (several months). This chapter investigates the interplay between social, material and cognitive dimensions of emerging computational and scientific literacies through the development of *sociomathematical norms* (McClain & Cobb, 2001; Yackel & Cobb, 1996; Cobb, Wood, Yackel, & McNeal, 1992). This chapter advances the argument that the teacher's emphasis on *mathematizing* and *measurement* as key forms of learning activities helped to meaningfully establish computation as the "language" of science in the classroom.

Following McClain, Cobb and colleagues (2001, 1996, 1992), this chapter defines sociomathematical norms as social norms that are uniquely mathematical, emerge through interaction with a mathematical object (such as a programming environment) and are given social value by the practicing community. Both in the work of Cobb and colleagues and our studies, the norms were often teacher-initiated. For example, the norms for assessing a computational model's "goodness" were largely based on the classroom teacher's personal conceptualizations of "accuracy", rather than being initiated by researchers and then taken up by teachers and students, (see for example, Lehrer, Schauble, Strom & Pligge (2001) and Manz (2012)). Such conceptual dissonances between the researchers and teacher might be construed as problematic, and certainly, from the perspective of disciplinarily accepted, canonical definitions and practices, they can be regarded as problematic. For example, the criteria for a measure to be statistically accurate are quite different from how the teacher in our study used the term. However, we believe that these dissonances are necessary to be studied and reported for understanding how new literacies are adopted and appropriated in classroom culture, and they can also inform potential areas of further collaboration between researchers and teachers.

In the study design, a third-grade teacher, in partnership with researchers, integrated agent-based programming with her regular science curriculum by iteratively developing sociomathematical norms for modeling motion using agent-based computational models. This chapter more deeply considers the interplay between diSessa's three pillars of literacy and examines the complex interplay and negotiations between modeling (material), reason (cognitive) and discourse and culture (social) involved *from the perspective of a teacher* with no prior background in programming or computational modeling. Throughout the year, the teacher emphasized connecting computational modeling to other out-of-computer modeling experiences, such as embodied and physical modeling activities, as well as re-framed computational representations as analogous to meaningful lived experiences for both herself and the students. The classroom teacher taught all lessons during this study and any changes to activities were made based on her formal and informal assessments of student understanding of the material or in-the-moment responses to student ideas. The teacher regularly initiated and supported the development of conventions for both "showing" and "knowing" during her instruction. The study investigates how the teacher adapted and employed this approach as a way to integrate computational modeling with her regular science curriculum to support the co-development of scientific and computational literacy in the classroom.

As a set, this work contributes to our understanding of how computational thinking and programming can transform education, in particular science education. Together, these papers illustrate how integration of computation as a language of science in the elementary classroom involves careful consideration of the complex interplay between materials, both computational and non-computational, cognition and classroom culture. Chapter 4, in particular, greatly extends current research on computing in the classroom by viewing integration through the eyes of the classroom teacher, highlighting the complex social dimensions that allow (or do not allow) various computational competencies to thrive in a classroom setting.

References

Blikstein, P. & Worsley, M. (2016). Children are not hackers. *Makeology: Makerspaces as learning environments*, *1*, 64.

diSessa, A. A. (1995). The many faces of a computational medium: Teaching the mathematics of motion. *Computers and Exploratory Learning*, 337-359.

diSessa, A. A. (2001). *Changing minds: Computers, learning, and literacy*. Mit Press.

Franz, G., & Papert, S. (1988). Computer as Material: Messing about with Time. *Teachers College Record*, *89*(3), 408-17.

Guzdial, M. (1994). Software-realized scaffolding to facilitate programming for science learning. *Interactive Learning Environments*, *4*(1), 001-044.

Harel, I. E., & Papert, S. E. (1991). *Constructionism*. Ablex Publishing.

Lehrer, R., Schauble, L., Strom, D., & Pligge, M. (2001). Similarity of form and substance: Modeling material kind. *Cognition and instruction: Twenty-five years of progress*, 39-74.

Maloy, R. W., & LaRoche, I. S. (2014). *We, the students and teachers: Teaching democratically in the history and social studies classroom*. SUNY Press.

NGSS Lead States. (2013). *Next generation science standards: For states, by states*. National Academies Press.

Papert, S. (1980). *Mindstorms: Children, computers, and powerful ideas*. Basic Books, Inc..

Papert, S. (1988). Does easy do it? Children, games, and learning. *Game developer magazine*.

Peppler, K., & Kafai, Y. (2005). Creative coding: Programming for personal expression. *Retrieved August*, *30*(2008), 314.

Schmidt, D. C. (2006). Model-driven engineering. *Computer IEEE – Computer Society. 39*(2), 25.

Sengupta, P., Dickes, A., Farris, A. V., Karan, A., Martin, D., & Wright, M. (2015). Programming in K-12 science classrooms. *Communications of the ACM*, *58*(11), 33-35.

Sengupta, P., Kinnebrew, J. S., Basu, S., Biswas, G., & Clark, D. (2013). Integrating computational thinking with K-12 science education using agent-based computation: A theoretical framework. *Education and Information Technologies*, *18*(2), 351-380.

Sengupta, P., & Shanahan, M. C. (In Press). Public Computation & Boundary Play: New Directions in STEM Education. *International Journal of Engineering Education.*

Sherin, B. L. (2001). How students understand physics equations. *Cognition and instruction*, *19*(4), 479-541.

Sherin, B., diSessa, A. A., & Hammer, D. (1993). Dynaturtle revisited: Learning physics through collaborative design of a computer model. *Interactive Learning Environments*, *3*(2), 91-118.

Street, B. V. (1984). *Literacy in theory and practice* (Vol. 9). Cambridge University Press.

Vallgårda, A., & Redström, J. (2007, April). Computational composites. In *Proceedings of the SIGCHI conference on Human factors in computing systems* (pp. 513-522). ACM.

Vee, A. (2013). Understanding computer programming as a literacy. *Literacy in Composition Studies*, *1*(2), 42-64.

Wilensky, U. (1995). Learning probability through building computational models.

Wing, J. M. (2006). Computational thinking. *Communications of the ACM*, *49*(3), 33-35.

Wing, J. M. (2008). Computational thinking and thinking about computing. *Philosophical transactions of the royal society of London A: mathematical, physical and engineering sciences*, *366*(1881), 3717-3725.

Whitehouse.gov (2016). *Computer Science for All*.

Yackel, E., & Cobb, P. (1996). Sociomathematical norms, argumentation, and autonomy in mathematics. *Journal for research in mathematics education*, 458-477.

CHAPTER II

LEARNING NATURAL SELECTION IN 4TH GRADE WITH MULTI-AGENT-BASED COMPUTATIONAL MODELS[1]

Introduction

Emergent phenomena are central in several domains such as biology (e.g., natural selection & evolution), economics (e.g., behavior of markets), and physics (e.g., statistical mechanics, thermodynamics & electromagnetism) (Darwin, 1989; Smith, 1977; Maxwell, 1871; Mitchell, 2009). Emergence is the process by which complex phenomena (e.g., flocking of birds, formation and collective movement of a traffic jam, formation of ant colonies, etc.) arise out of the interactions between many individual objects, actors or agents (Strogatz, 2003; Holland, 1998; Wilensky & Resnick, 1999; Resnick, 1994). The formation and overall movement of a traffic jam is an example of an emergent phenomenon that we experience frequently in our daily lives: A traffic jam results from individual cars speeding up and slowing down, even when individual cars do not intend to cause the jam. And, while the individual cars move forward, the overall jam propagates in a backward direction due to the aggregation of delays between cars stopping and starting (Resnick, 1994).

Complex systems educators have shown that developing a deep understanding of emergent phenomena involves being able to develop multi-level or agent-aggregate-complementary explanations – i.e., explanations that involve causal relationships between the individual level agents' attributes, behaviors and interactions on one hand, and the collective

---

[1] This chapter was published in *Research in Science Education* in 2013.

aggregate-level behavior on the other (Resnick, 1994; Chi & Ferarri, 1998; Jacobson & Wilensky, 2006; Hmelo-Silver, Marathe & Liu, 2007; Wilensky & Resnick, 1999; Penner, 2000; Sengupta & Wilensky, 2009, 2011). However, such emergent behaviors have been found to be counter-intuitive for novices, making it challenging for them to understand in absence of suitable instructional supports (Wilensky & Resnick, 1999; Sengupta & Wilensky, 2009; Jacobson & Wilensky, 2006).

Natural selection, which is a principal mechanism of evolution, is also an emergent phenomenon (Wilensky & Reisman, 2006; Chi, 2005; Gould, 1996). Researchers have shown that students at all levels (middle school through college) find natural selection a very challenging topic to understand. They have suggested two primary reasons for this difficulty: a) From a cognitive perspective, middle and high school students often bring in alternative explanations and theories regarding natural selection that are resistant to change through instruction (Chi & Ferrari, 1998; Chi, 2005; Anderson, 2002; White, 1997); and b) From an instructional perspective, understanding natural selection is dependent on understanding complex emergent processes that are not easily visualized in traditional classroom instruction, and therefore difficult to the student (Dodick 2003, Inagaki & Hatano 2002; Wilensky & Reisman, 2006). We discuss both these perspectives below.

From a cognitive perspective, research in children's early biological knowledge suggests that children draw analogically on their knowledge about humans when presented with unfamiliar biological phenomena (Piaget, 1929; Carey, 1985; Inagaki & Hatano, 2002). Children possess fairly rich knowledge about their own biological needs (as humans) in comparison to other plants and animals, and their biological reasoning is often teleological in nature (Coley, 2000; Keil, 1992; Kelemen, 1999; Solomon, 1996). For example, Inagaki and Hatano (2002) found that pre-school children believe that organs and other biological functions have agency and purpose for the actions they perform. However, although pre-school and elementary children have such a repertoire of biological knowledge based on which they may construct new inferences about biological phenomena, causal explanations regarding more complex phenomena such as population growth, reproduction, death, decline and inheritance of traits - which are essential epistemic components for understanding natural selection - are difficult for learners to develop and understand even at the secondary level (Hendrix, 1981; Bernstein, 1975; Inagaki and Hatano, 2002).

Chi and her colleagues further argued that novice's alternative ideas about natural selection result due to an ontological miscategorization of natural selection as a direct or an event-like process rather than an equilibration process (Chi & Ferrari, 1998; Chi, 2005). An event process is one that is distinct, sequential and bounded, having a clear beginning and end and often a specific goal. In contrast, an equilibration process is one that is uniform, simultaneous and ongoing with no defined beginning or end. Natural selection, in this view, is an example of an ongoing, equilibration process, and the underlying reason for students' misconceptions is that they miscategorize natural selection as a sequential and bounded process (Chi & Ferrari, 1998; Chi, 2005).

From an instructional perspective, several scholars have argued that in most science classrooms, aggregate-level formalisms are typically used to teach population dynamics in a predator-prey ecosystem. An example is the Lotka-Volterra differential equation, which depict how populations of different species in a predator-prey ecosystem evolve over time. This equation still forms the cornerstone of classroom instruction in natural selection at the high school level or beyond (Wilensky & Reisman, 2006). While mathematically correct, these formalisms do not make explicit the underlying agent-level attributes and interactions of the system, and as such remain out of reach of younger students (e.g., elementary school students). On the other hand, in the context of understanding emergent phenomena, agent-based reasoning, i.e., reasoning about the attributes and behaviors of the individual, is developmentally prior to aggregate reasoning (Levy & Wilensky, 2008; Sengupta & Wilensky, 2009, 2011; Blikstein & Wilensky, 2009). This is because agent-level reasoning leverages children's intuitive knowledge about their own bodies, perceptions, decisions and actions (Papert, 1980; diSessa, 2000; Levy & Wilensky, 2008). This body of research suggests that it is the lack of connection between students' natural, embodied agent-based reasoning on one hand, and the aggregate forms of reasoning and representations they encounter in school on the other, that creates a barrier to their understanding of emergent phenomena.

Several scholars have also shown that modeling-based curricula based on multi-agent-based models can address this divide, as these models (and the learning activities designed around them) help novices understand emergence by recruiting their agent-level intuitions (Resnick, 1994; Wilensky & Resnick, 1999; Klopfer, Yoon & Perry, 2005; Klopfer, Yoon & Um, 2005; Sengupta & Wilensky, 2009, 2011; Blikstein & Wilensky, 2009; Wilensky, 2003). The term "agent" in the context of multi-agent-based models (hereon referred to as MABMs), indicates individual computational objects or actors (e.g., cars, in the traffic jam example cited earlier), which obey simple rules assigned or controlled by the user. It is the interactions between these agents (based on the rules assigned or controlled by the user) that give rise to collective, aggregate-level behavior (e.g., formation and movement of the traffic jam). A pedagogical approach based on MABMs, as Sengupta & Wilensky (2009) pointed out, emphasizes the continuity between novices' pre-instructional ideas (many of which have traditionally been regarded as misconceptions) on one hand, and the development of deep, multi-level, expert-like conceptualization of emergent phenomena on the other. For example, in analyzing how students learn electrical conduction by using MABMs, Sengupta & Wilensky (2011) found that learners began with intuitive, non-canonical initial interpretations of salient elements of and events depicted in the models, and that these initial ideas (termed registrations) played an essential and productive role in the development of students' understanding of the relevant aggregate-level phenomena.

Our study is based on this approach. Whereas prior research has shown that high school and college students can develop a deep understanding of population dynamics by building MABMs (Wilensky & Reisman, 2006), in this paper, we investigate how elementary school students with no prior instruction in ecological systems or evolution, develop multi-level understandings of some introductory aspects of population dynamics in a simple predator-prey ecosystem, through scaffolded interactions with a MABM. Because MABMs leverage students' intuitive reasoning at the agent-level, we expect that students would use their intuitive ideas in order to interpret and explain relevant phenomena as they interact with the simulation. Our central research goals for this study include identifying students' intuitive knowledge in the form of their initial interpretations of the relevant phenomena, as well as the roles these initial ideas play in the process of their conceptual development in course of their scaffolded interactions with the model. In doing so, our goal is therefore to identify the process of bootstrapping a students' intuitive knowledge, as opposed to focusing how students' intuitive knowledge can be replaced with canonically correct knowledge. Thus, our approach is grounded in the constructivist approach for understanding conceptual change in students (Smith, diSessa &

Roschelle, 1994; Hammer, 1996), in which expert-like understandings (concepts) can emerge from a recombination of naive knowledge elements.

In this paper, we report a semi-clinical interview based study conducted with 4th grade students in the domain of ecology, with a particular focus on population dynamics. In our study, we focused on the following: a) identifying the nature of learners' initial interpretations of salient events or elements of the represented phenomena, b) identifying the roles these interpretations play in the development of their multi-level explanations, and c) how attending to different levels of the relevant phenomena can make explicit different mechanisms to the learners. In addition, our analysis also shows that although there were differences between high- and low-performing students (in terms of being able to explain population-level behaviors) in the pre-test, these differences disappeared in the post-test.

Prior Research on Knowledge Analysis in MABM-Based Learning Environments

Because the characterization of students' knowledge is central to the purpose of this paper, it is important to discuss how researchers grounded in the constructivist perspective have investigated the process of conceptual development of learners' understanding of complex scientific phenomena using MABMs. The categories we developed for knowledge analysis in this paper (described in the following sections) are based on our understanding of the categories for knowledge analysis developed and used by these researchers.

Overall, in their analysis of the process of students' conceptual development, researchers have focused on two dimensions: a) the "levels" of learners' explanations (Abrahamson, 2004; Blikstein & Wilensky, 2009; Wilensky & Resnick, 1999; Sengupta & Wilensky, 2009; Jacobson & Wilensky, 2006; Levy & Wilensky, 2008), and b) form and structure of knowledge elements used by learners in the process of knowledge construction (Sengupta & Wilensky, 2010, 2009).

Along the first dimension, the majority of the research on naïve cognition of complex systems using MABMs have focused on two *levels of description* and two associated *modes of reasoning* that students and experts utilize when trying to explain complex systems (Abrahamson, 2004; Blikstein & Wilensky, 2009; Wilensky & Resnick, 1999; Sengupta & Wilensky, 2009; Jacobson & Wilensky, 2006). The two levels are the micro and macro levels: the micro level involves the behavior of individuals or agents, and the macro level relates to the group properties. The corresponding modes of reasoning are agent-based and aggregate reasoning. Specifically in the context of ecology and evolution, Wilensky & Reisman (2006) and Wilensky & Stroup (2003) showed that novice learners (high school and undergraduate students) were able to model aggregate-level behaviors of complex biological phenomena, such as population dynamics in a predator-prey ecosystem and spread of diseases, by first identifying and specifying agent-level rules and interactions. That is, students first engaged in agent-level reasoning, and through modeling activities that involved being able to identify (and iteratively, refine) the effect of these agent-level rules on the aggregate-level outcomes, they were able to generate correct explanations of the aggregate-level phenomena. These explanations, although qualitative in nature, were consistent with aggregate-level equation based representations (e.g., Lotka-Volterra equations), and involved correct identification of relevant agent-level properties as well. Wilensky & Stroup (2003) termed such explanations *agent-aggregate complementary.*

Along the second dimension, our particular focus for this paper is on the nature of learners' initial interpretations of the phenomena depicted in the MABMs. Pertaining to this issue, there currently exists no research on learning biology using MABMs. However, Sengupta and colleagues have focused on identifying the form of initial explanations and interpretations of novice learners, and their role in the learners' process of knowledge construction as they interact with MABMs, in the domain of electricity. Similar to ecosystems, electrical conduction, as Sengupta & Wilensky (2009) argued, can also be represented in as a complex phenomenon arising from interactions between many microscopic individual agents (e.g., electrons and ions). Sengupta's research shows that when middle school students (5th and 7th graders) construct their understandings of electrical conduction through interacting with MABMs, they use the following types of knowledge structures: Registrations and Causal Schemas (Sengupta & Wilensky, 2010, under review). Based on Roschelle (1991), they defined registrations as the elements depicted in an MABM model that become salient to the learner during the course of an observation. Registrations indicate how learners *initially* parse the phenomena represented in the models. A registration thus provides a learner with an initial "framing" of what's going on the model. Sengupta & Wilensky (2011) defined *causal schemas* as binary relations among variables that learners notice in the model, typically in the form of "A causes B". Causal schemas can in some cases be weak forms of explanations, since the learner may not always elaborate the mechanisms represented in causal schemas. Sengupta and Wilensky (2008) found that learners develop more complex causal mechanisms by coordinating multiple causal schemas.

Sengupta and Wilensky (2008, 2011) found that learners' initial registrations played an important role in the overall process of knowledge construction in the learners during their later interactions with MABM(s). For example, they found that students who initially interpreted the process of electron flow as translational movement were able to easily identify the individual effects of the number of charged particles, and their speed, on the rate of overall flow. In contrast, when students initially interpreted the process of electron movement as a process of accumulation inside the battery terminal, they were able to identify the compensatory effect of both the number and speed of the charged particles on the rate of flow, and thus develop a more complex causal schema and explanation for electron flow.

Research Questions

In this study, we investigated the following research questions:

1. What is the process of conceptual development that takes place during learners' interactions with the model? Specifically, we investigated two issues:
   a. What is the role of learners' initial interpretations of the depicted phenomena in the development of their multi-level explanations?
   b. How do learners develop multi-level explanations of the phenomena displayed in the simulation by paying attention to different levels of the phenomena?
2. What are the qualitative differences between students' pre- and post-test responses? Specifically, we investigated two issues:
   a. What are the differences in terms of multi-level structure of students' pre- and post-test explanations?
   b. Do high and low performing students show evidence of learning gains in terms of their ability to reason about the pre- and post-test questions from an agent/aggregate complementary perspective?

The Learning Environment: Bird-Butterfly Random Phenotype Model

The model used in this study, Birds & Butterflies Random Phenotype Model (in short: Birds & Butterflies Model), was designed in the NetLogo modeling platform (Wilensky, 1999). NetLogo is a multi-agent-based programming language and modeling platform widely used by educators and scientists for designing multi-agent based models of scientific phenomena. The notion of an "agent" in NetLogo, in turn, draws inspiration from Papert's protean Logo turtle

(Papert, 1980). The Logo turtle is a computational object, using which very young children can successfully learn mathematical concepts and phenomena by constructing behaviors of the turtle that are “body-syntonic” (Papert, 1980). Papert called body-syntonic the use of tools merging one's own bodily understanding with formal (mathematical) ideas – e.g., a drawing a circle using a Logo turtle involves repeating “body-syntonic” actions such as moving forward and then turning by a small amount (Papert, 1980).

In a NetLogo model, learners (or users) can create or control thousands of Logo turtles (agents), and create or modify simple rules for each agent (or groups of agents) to obey. It is through the interaction between these agents based on these rules that emergent, aggregate-level phenomena arise.

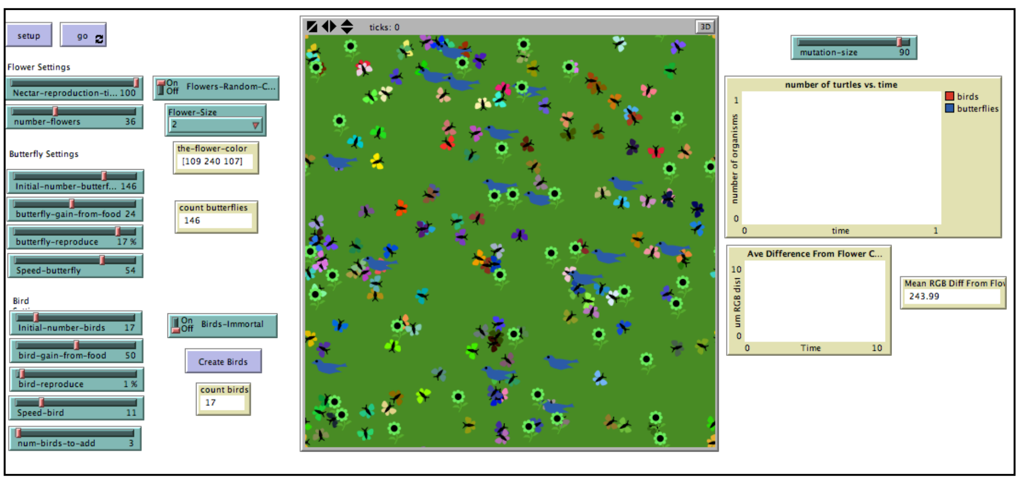


Figure 1: Screenshot of User Interface of the Birds & Butterflies Model

Our model consists of three “breeds” of agents: birds that act as predators, butterflies that act as prey, and flowers that constitute the natural environment. Each member of these three breeds obeys a set of simple rules: birds prey on butterflies, butterflies survive by drinking nectar from the flowers (and gain energy in the process), and butterflies whose colors are closer to that of the flowers are harder to be seen by the birds. When the simulation is run over several iterations, it shows that butterflies change colors over multiple generations and gradually become similar in appearance to the flowers. This is because the butterflies with colors closer to that of the flowers survive predation and thus get naturally selected every generation. A detailed list of the agent-level rules is listed below in Table I along with the corresponding aggregate behavior(s).

The aggregate-level phenomena are illustrated both by change in attributes and/or behavior of individual agents (e.g., change in color of a particular species over successive generations; change in total number of agent's particular species), as well as represented in the form of graphs. Rather than simply learning the vocabulary, students are therefore able to interact with the model through altering individual-level rules and variables (e.g., energy gain from food for both birds and butterflies; initial color and number; etc.), which in turn leads to, or affects natural selection. These agent-level behaviors are intuitive, and therefore easily recognizable by the learners (e.g., animals need food to survive; camouflage; etc.). The ability to thus 'see' natural selection emerges in real time over successive generations of butterflies and under different environmental constraints from aggregations of these simple rules of interaction, we hypothesize, will deepen students' understanding of natural selection, without necessitating them to discard their intuitive knowledge.

Table I: Individual level rules and corresponding emergent outcomes in the Birds & Butterflies Random Phenotype Model

Agent	Rules	Resulting Aggregate Behavior
Birds	1. Birds need energy to survive, they gain energy from eating butterflies and lose energy by flying and reproducing 2. If the Birds' energy falls below a certain level, they die. 3. Birds randomly reproduce and will continue to reproduce if their energy is high enough 4. To catch butterflies, birds are instructed to be more attracted to (better able to 'see') the butterflies the furthest away (in the color spectrum) from the color of the flowers 5. Birds will consume the nearest attractive butterfly 6. Bird color is uniform as is energy gained from eating butterflies.	1. Birds selectively favor butterflies of certain colors – specifically those the least like the colors of the flowers 2. Butterflies unsuited to their environment (based on color) are at a disadvantage and stand a greater chance of being eaten
Butterflies	1. Butterflies need energy to survive, they gain energy by drinking nectar from the flowers and lose energy by fluttering and reproducing 2. If the Butterfly's energy falls below a certain level, they die 3. Butterflies randomly reproduce and will continue to reproduce if their energy level is high enough 4. Butterflies are programmed to gain more energy from flowers that are closer in color to themselves 5. Butterfly color is random 6. Butterflies reproduce like offspring	1. The environment (flower color) selects for certain butterflies to be more successful than others 2. Butterflies closer in color to the flowers gain more energy and are reproductively more successful 3. There is differential survival and reproductive success of butterflies with certain traits
Flowers	1. Flower color is uniform but will reset to a different uniform color with each restart of the program 2. Flowers have equal amounts of nectar and will reproduce nectar after a butterfly has visited the flower after a uniform countdown	1. Flower color results in certain butterfly traits, namely color, to be more successful than others.

Method

Setting, Study Design & Procedure

The setting of this study was a 96% African-American charter elementary school located in a metropolitan school district of a Southeastern state in the US. We adopted a contrasting groups design with both qualitative interviews and quantitative analysis. Pullout clinical interviews were conducted with ten 4^{th} grade students of mixed gender representing the top five and bottom five performers in the class. Students in each group (high- and low-performing) were selected by the 4^{th} grade science teacher based on their comprehensive academic performance in science throughout the year. On average, students spent roughly half an hour interacting with the model, preceded by a pre-test and followed a post-test. Overall, the study lasted about an hour for each student.

During the course of the study, students were first asked to answer three pre-test questions that focused on their prior understanding of population dynamics and change of over time. These questions are shown in Table III, and students were asked to answer these questions again in the post-test (after their interactions with the model). Student responses to pre- and post-tests were a mixture of both written and verbal responses. Verbal responses were elicited in order to clarify their written responses, wherever necessary. For instance, during the pre- and post-tests some students chose to write down their answers to questions almost immediately after reading the question – i.e., without any prompt from the interviewer. This elicited further clarifying questions from the interviewer *after* they had written down their responses. Other students chose to first state their answer aloud before writing anything down, thus prompting clarifying questions *both before and after* students had provided written answers. The primary goal of these questions was to prompt for mechanistic explanations to gain a deeper insight to the learners' conceptual understanding. For example, if a student indicated that population of a particular species would change, the interviewer would ask them to explain why they thought so (or what made them think so).

Upon completion of the pre-test, then began their interaction with computer model. A summary of the verbal protocol used during the course of the model simulations is provided below in Table II.

Table II: Summary of verbal prompts used during student interaction with the model

Activity	Time (min)	Verbal Prompts (with appropriate clarifying questions used throughout the interview)
Introduction	0-5	Interviewer explanation of the interface. Buttons such as "Setup" are introduced, Reporters (Count Birds, etc.) are pointed out and Sliders are explained.
Introduction	5-7	Student is asked to: 1) press the "Setup" button and 2) Describe what they see on the screen (the screen is static at this point in the interview)
Activity 1	7-10	Student is asked to press the "Go" button and describe what they see happening on the screen
Activity 2	10-13	Student is asked to hit the "Setup" button again and is asked to state what color the flowers are. The student is then asked to predict what they think will happen to the butterflies once they hit the "Go" button.
Activity 3.1	13-20	The student is asked to hit the "Go" button and observe what they see on the screen. The interviewer then asks the student to explain why the butterflies changed color and why the total population changed.
Activity 3.2	20-30	The student is asked to reset the program and press "Go". For this round of simulations, the student is asked to explain where the new butterflies came from and to predict/explain what color the new butterflies would be.

During this phase of the study, the interview prompts played a dual role. Besides prompting students for mechanistic explanations of their observations and predictions, these prompts were also designed to act as scaffolds, by directing students' attention to particular aspects of the model. Some of these scaffolds introduced students to relevant elements and their functions (necessary for manipulating the model) in the user-interface of the model. For example, students were introduced to the model through a brief (approximately 3-5 minute) oral introduction, where the interviewer explained the functions of the buttons ("Setup" and "Go"), sliders (variables), and reporters ("Count bird" and "Count butterfly", that displayed the instantaneous total number of birds and butterflies, respectively). Students were then asked to setup the model with run the model with the initial settings and describe what they were observing on the screen (i.e., what they thought was going on in the model). In asking this question, our goal was to gain an understanding of students' initial interpretations of the phenomena depicted in the model, or *registrations* (explained in the following section). Upon completion of the first simulation run, students were next asked to reset the program and through directed verbal prompts, the students' attention was drawn to an important variable: the flower color. Flower color was intentionally programmed to randomly change each time the model was "reset", to help students observe the direct effect of environmental selective pressure on the behavior of the individual agents as well as aggregate-level outcomes such as population change. Students were then prompted to make predictions about the aggregate-level behaviors of the birds and butterflies, based on their observations in the previous simulation. Note that the sequence of these prompts (in terms of the levels of phenomena that students were requested to direct their attention to) was designed based on prior research on MABMs discussed earlier in this paper, which shows that students' agent-level reasoning is developmentally prior to aggregate-level reasoning. These prompts, therefore, besides serving as interview prompts, also acted as scaffolds to aid the process of conceptual development in students. Students were then asked to run the program to test their predictions and offer explanations of what they were observing. This cycle of prediction-simulation-explanation was repeated two to three additional times with each subsequent trial focusing on aggregate-level phenomena such as population change over time, and their underlying causes.

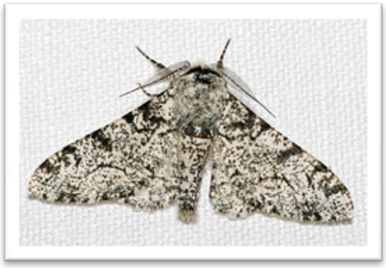
Pre- and Post-Tests

The questions in the pre-test and post-test were designed to assess student thinking both at the individual level (see questions 1 and 2 in Table III) and at the aggregate-level (see question 3 in Table III). In both the pre- and post-test, students were first asked to read the scenario described in Figure 2, and then answer the questions listed in Table III.

Table III: Sample Pre-/Post-Test Questions

Question 1	Which type of tree trunk would the light-colored moth prefer to rest on - the lichen covered tree trunk, or the lichen free tree trunk? Why?
Question 2	If there is both a light-colored moth and dark colored moth on a lichen free tree trunk, which moth is the Great Tit more likely to eat and why?
Question 3	How would the moth population change if all of the lichen were to die and none of the tree trunks were covered in lichen?

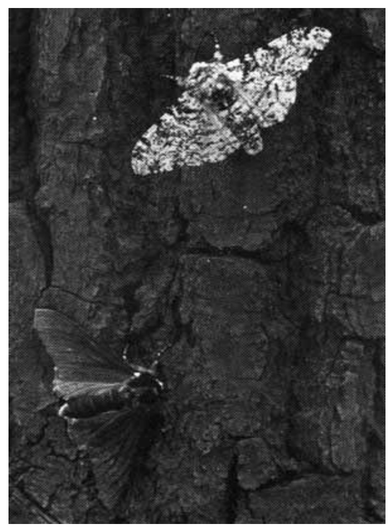
A species of moth called the Peppered Moth exists in two color variations – light and dark (see pictures below).


Light Variation


Dark Variation

The Peppered Moth prefers to rest on the trunks of trees. Some tree trunks are covered in an organism called lichen that gives the tree trunk a lighter color, while other tree trunks do not have any lichen and are dark in color. Below are pictures of two moths on a lichen covered tree trunk and a lichen free tree trunk.


Moths Resting on a Lichen Covered Trunk


Moths Resting on a Lichen Free Trunk

The Peppered Moth is a common food source of the Great Tit (see picture below), a predatory bird that will eat all of the moths it can find.


Great Tit

Figure 2: Sample Pre-/Post-Test Scenario

Case Study Approach

In order to investigate the process of conceptual development during learners' interactions with the model, we adopted an explanatory case study approach for our analysis (Taber, 2008; Simons, 1980; Gomm, Hammersley & Foster, 2000; Yin, 1994). Case studies can provide us with a deeper understanding of an individual's point of view (Denzin & Lincoln, 2000). As Yin (1994) pointed out, while case studies may be of several types, *explanatory* case studies are well suited as a methodology to answer *how* and *why* questions. Because our central goal is to characterize the process of knowledge construction in the learners' minds as they interact with the simulation – i.e., a *how* question - we believe that an explanatory case study based approach is well suited for our research goals. The cases we highlight in the next section reveal the role that learners' intuitive knowledge play in their processes of knowledge construction, as they interact with the MABM.

We now explain the rationale behind the selection of our cases. Following Petri and Niedderer (1998), Taber (2008) presented two criteria for selection of cases: representativeness and typicality. Petri & Neidderer (1998), for example, chose a representative case in the form of a student who was present during every class and frequently interviewed, and thus aptly represented the entire instructional process. The criterion of typicality implies that the selected case should potentially offer insights, which are likely to have wider relevance. In this paper, for each learning activity, we discuss two cases in detail: Conitra and Larry. We chose Conitra because she is representative of the high performing students, as evident in her pre-test responses, and we chose Larry because he is representative of the low-performing students, as also evident in his pre-test responses.

We conducted both *within-case* and *between-cases* analyses. Pertaining to each learning activity, we first present the analysis of Conitra and Larry as discrete cases, and then present between-cases analyses by summarizing the findings from responses of all the participants. During the within-case analyses, based on the coding scheme presented in the next section, we identified the different types of knowledge structures that become evident in their verbal reports. All the participants in our study underwent the same sequence of activities, and this regularity provided us with the basis for comparison of students' verbal reports across the entire sample. Between-cases comparison enabled us to address the issue of typicality of individual cases, by helping us identify how prevalent particular forms of explanations were among all the participants, or in some cases, helping us identify the variations among participants, e.g., the different instantiations of the same knowledge structure, or different knowledge structures being used by different participants in the same context.

Coding Scheme

Our coding scheme is theoretically grounded in prior research on knowledge analysis in MABM based learning environments, which we have discussed earlier in this paper in the section on knowledge analysis in MABM-Based learning environments. We categorized learners' responses in terms of the following forms of knowledge structures reflected in their responses: registrations, causal schemas, and levels-based perspectives. We discuss next the definition of each category, along with some sample responses and explanations of how we identified them, and this discussion is also presented in a summarized form in Table IV.

Table IV: Category Types and Identification Heuristics

Category	Heuristics for Identification	Sample Responses
Registrations	*Operational Definition:* Registrations typically indicate student's first interpretive observations of salient elements displayed in the model during a run.	• *Changing color:* "the butterflies are changing color" • *Piling Up:* "the butterflies are piling up"
Causal Schemas	*Operational Definition:* Causal Schemas are model specific cause-and-effect relationships that typically take the form "A causes B" in students' responses. [We denote causal schemas in the text as A → B]	• "The population will go down since those butterflies have no more camouflage." (*Causal Schema: Camouflage → Survival*)
Levels Based Perspectives	*Operational Definitions:* *Agent Perspective*: Students' responses mention agent-level attributes and behaviors. *Aggregate Perspective:* Students' responses mention aggregate-level attributes and behaviors *Agent-Aggregate Complementarity:* student responses explain aggregate-level outcomes using the agent perspective.	• *Agent Perspective:* "The light moth will rest on the lichen because it is soft." • *Aggregate Perspective:* The population will get big, big" • *Agent-Aggregate Complementarity Perspective:* "The dark moth population will go up because they will have babies because they're not eaten."
Agent-Aggregate Links (AALs)	*Operational Definition:* Agent-Aggregate links represent the number of distinct agent-level attributes (including behaviors) that students use to explain an aggregate-level phenomenon.	• "The population will start going down, the light-colored moths will go down because they are easier to see so they might get eaten." *Analysis:* AAL1: "moths get eaten (by birds)", and AAL2: "(moths) can be seen by the birds". In the case of AAL1, the relevant agent-level behavior used is "eating", and in the case of AAL2, the relevant agent-behavior used is "seeing". Both AALs are being used to explain an aggregate-level phenomenon: "population will start going down"

Initial interpretations or registrations. As discussed in an earlier section, following Roschelle (1991) and Sengupta & Wilensky (2011), we define registrations as the elements depicted in an MABM model that become salient to the learner during the course of an observation. A registration thus provides a learner with an initial "framing" of what's going on the model. In order to identify students' initial interpretations of events or elements (depicted in the model) that appeared salient to them, we asked the following question at the beginning of the interview session. Typically, registrations became evident in students' responses when we first introduced students to the basics of the simulation interface, and asked them the following question: "Tell me what you think is going on in the model". Students' responses to this question revealed the elements of the simulation that were salient to them, and often varied in terms of the interpretiveness. For example, in Activity 1, the movement of the birds and the butterflies on the screen registered in the students' minds as "the butterflies are changing color" (i.e., the focus here is on camouflage) in some cases and in other cases as "the butterflies are piling up" (i.e., the focus here is on the increase in population of butterflies on the screen).

Causal Schemas. Sengupta & Wilensky (2008, 2011) defined causal schemas as binary relations among variables that learners notice in the model, typically in the form of "A causes B". Causal schemas can in some cases be weak forms of explanations, since the learner may not always elaborate the mechanisms represented in causal schemas. We identified causal schemas by analyzing students' responses when we asked them to explain their observations or predictions during Activity 3. Causal schemas were evident in students' responses when we asked them to explain the effects of changing a particular variable - such as the flower color, or the color of the butterflies – on the behavior of the overall simulation. For instance, when a student was asked to explain why the population of a particular color of butterflies was increasing, he replied: “The population will go down since those butterflies have no more camouflage.” This response is premised on the observation that the butterflies were changing color to blend in with the flowers so as to not be eaten, and is indicative of the causal schema that camouflage leads to survival.

Agent-aggregate complementarity (or Multi-Level Responses) in pre- and post-test. In order to answer our second research question, we investigated the multi-level nature of students’ responses in the pre- and post-test. We coded their responses according to their use of an agent-perspective, aggregate-perspective and/or an agent/aggregate complementary perspective. Consider, for example, the following responses:

R1: “The light-colored moth would get caught.”

R2: “The population will get big, big.”

R3: “The light moths will get eaten because they can be seen by the birds”

As explained by Sengupta and Wilensky (2009), an aggregate-only perspective is an explanation that is devoid of any mention of micro-level agents and/or interactions. An example of this is R2, where the student only explicitly mentions an aggregate-level outcome. Conversely, an agent-perspective would indicate an explanation that involves explicit mention of the individual-level agents and their interactions, without any explicit mention of an aggregate-perspective. An example of this is R1, where the students’ response in concerned with the behavior of an individual moth. Lastly, a complementary perspective would indicate an explanation that in addition to explicitly mentioning the agent-perspective, also describes how the aggregate-level phenomena emerges from the agent-perspective. An example of this is R3, where the change in aggregate-level behaviors is explained by using interactions between agents. It is noteworthy that we were unable to find agent-level only explanations in students’ written pre- and post-test responses to Q3. This was due to the nature of the question that students were asked to answer, which explicitly prompted them to explain an aggregate-level phenomenon.

To develop a quantitative measure for assessing agent-aggregate complementarity, we coded students' responses in terms of the number of distinct agent-level attributes (including behaviors) that they used to explain an aggregate-level phenomenon. We refer to each agent-attribute (when used to explain an aggregate-level phenomenon) as an Agent-Aggregate Link (AAL). For example, the statement R3 quoted above was used by the student to explain why the population of light moths would decrease. Note that this statement therefore has two AALs: AAL1: "moths get eaten (by birds)", and AAL2: "(moths) can be seen by the birds". In the case of AAL1, the relevant agent-level behavior used is "eating", and in the case of AAL2, the relevant agent-behavior used is "seeing".

Reliability

The codes for qualitative analysis of the interview data as well as the written pre- and post-test data were developed through mutual discussion between the two authors. Both the authors jointly coded the data over a period of several months, and through regular discussion, grounded in the theoretical framework discussed previously, came to agreement about application of the codes. Once agreement between the two authors was reached, all of the pre- and post-test data were blind-coded by Levy, who was provided with the coding scheme. Levy is a researcher who is a member of our research lab but was not involved with the study and had not seen the data previously. Levy's codes agreed with the authors' codes 95% of the time, resulting in a Cohen's Kappa of 0.9. In addition, a second coder, Ravit, independently coded 20% of the interview data, and agreed with the authors' codes 95% of the time, resulting in a Cohen's Kappa of 0.9. The second coder is not a member of our research lab, and is a qualitative researcher in the social sciences. She was also not affiliated with the study and had not seen the data previously.

Findings

The Development of Students' Understanding

Activity 1: Students' initial interpretations. The goal of the first activity was to elicit students' initial interpretations (registrations) of what is going on in the model. Each registration, as discussed earlier, is an interpretive action by the learner, in which he or she identifies the elements within the model that appear salient to them, and his or her interpretation of the salience of those elements. The registrations we identified varied in the degree of interpretiveness, i.e., some registrations indicated events that were directly seen by the students (e.g., changing color; piling up), whereas in some other cases, registrations involved students ascribing intentional acts to agents (e.g., tricking). To better illustrate how students' initial interpretations served as a platform from which they developed their later explanations, we will follow the cases of two students in the rest of this section: Conitra, a high performing student and Larry, a low performing student.

Conitra's Case. Conitra's initial explanations of the behavior of the agents within the system were directly based on her initial observations. As the following excerpt shows, during Conitra's initial interaction with model, she first noticed that 1) the butterflies were changing color (line 3), and 2) that more butterflies were appearing on screen (line 5). After continuing to observe the model, the interviewer then asked Conitra if she noticed anything more specific about the color of the butterflies to which Conitra specified that butterflies were changing to the same color as the flowers (line 8).

Excerpt 1

Interviewer: What's happening? Can you see any changes with the butterflies?

Conitra: It look like they *changing colors*.

Interviewer: Okay, they're changing colors.

Conitra: There's *more butterflies*...

Interviewer: Uh huh, yeah we have more butterflies. And do you notice anything special about the color that they're changing to?

Conitra: They *changing to the same color as the flower*...

When the interviewer asked Conitra to run the simulation again (see Excerpt 2), Conitra refined her initial observation regarding the increased population of the butterflies by articulating that the butterfly population had in fact fluctuated before steadily increasing to a higher number. Conitra builds on her initial observation that the butterfly population had increased by observing that the population had in fact first increased, then decreased and then increased again (lines 2 and 3) before eventually stabilizing at around 300 butterflies (line 8). Here the events that became salient to Conitra are as follows: a) fluctuation in the butterfly population, with an eventual increase of the population, and b) that the butterflies (that survived) changed color to the same as that of the flower. We term (a) the *fluctuation* registration, and (b) the *changing color* registration. As we will see in Activity 3, this observation, combined with her initial observations regarding the butterflies changing color, provided a platform from which all of Conitra's later explanations were constructed.

Excerpt 2

Interviewer: Okay, go ahead and hit go and let's see what happens.

Conitra: They *went up, and then they went down, and then they went up.*

Interviewer: Yeah, now they're going back up again. And so, do you remember how many butterflies we started with?

Conitra: 146

Interviewer: Right, and how many do we have right now?

Conitra: [Conitra *states several numbers around 300* as the population continues to fluctuate]

Interviewer: Yeah, it's around 300. Okay, and what happened to the butterfly population?

Conitra: It increased.

Interviewer: It did, it increased, what else happened to the butterfly population?

Conitra: It changed.

Interviewer: It changed, in what way did it change?

Conitra: Color.

Larry's Case. Larry offers an interesting contrast to Conitra's observations described above. In his initial interaction with the model, Larry also notices that the butterfly population was increasing, however, his initial explanation involves a greater degree of interpretiveness than Conitra's. Larry interprets the increase in population as the butterflies piling on top of each other, and we have termed this interpretive observation the "Piling Up" registration.

Excerpt 3

Interviewer: What do you see happening on the screen? Especially with the butterflies, what's happening with the butterflies?

Larry: They're moving around.

4 Interviewer: Yes, they're moving around. What else do you notice is
5 different?
6 Larry: They're leaving and the butterflies are *just standing in a pile*
7 and the birds flying by.

In the above excerpt, Larry initially observes many more like-colored butterflies flying around the flowers. The crowded image depicted on the screen is viewed as the butterflies 'piling up' on top of each other (line 6) rather than as the butterflies substantially increasing in population. A screenshot corresponding to Larry's mention of the 'piling up' registration is provided below in Figure 3.

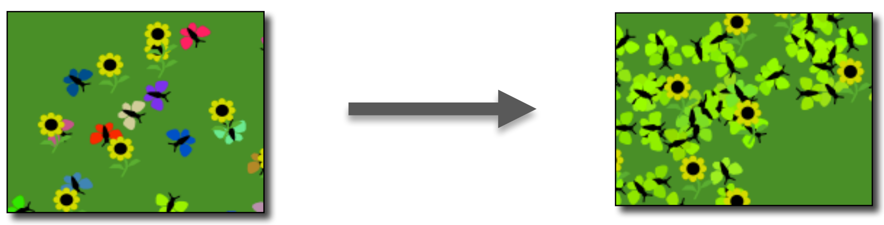


Figure 3: A Representation of the 'Piling Up' Registration

Another particularly interesting registration invoked by Larry is the *tricking* registration. In the following excerpt, in which the interviewer asks Larry which butterflies he thinks might not get eaten when he starts the simulation, the following exchange ensued:

Excerpt 4

1 Interviewer: Why would those butterflies get eaten?
2 Larry: They have lighter colors.
3 Interviewer: Okay, because they are lighter than others [butterflies].
4 Which ones might not get eaten?
5 Larry: The purple and blue ones.

Interviewer: Why do you think those might not get eaten?

Larry: Because they look just like the birds.

Interviewer: Why do you think looking like the birds might help them not get eaten?

Larry: *Because the bird might think that's it's its cousin*.

Interviewer: Okay, so they're going to think that they're, like, family - that they're one of them. Um, okay, well why don't you hit go and let's see which butterflies don't get eaten.

In excerpt 4, Larry is explaining his prediction that the lighter color butterflies (i.e., purple and blue butterflies) might not get eaten (lines 1 - 5) because they are similar to the color of the birds (line 7). To explain his prediction, Larry suggests that the birds might not eat them because they might think that the blue and purple butterflies are their *cousins* (line 10). We term this registration the *tricking* registration; because it is premised on the idea that birds might be *tricked* into thinking that the butterflies should not be eaten. Although Larry’s explanations are more interpretative than Conitra’s, his initial ideas are based on his observation of the same behaviors as Conitra: changing color and population growth - both of which served as a platform from which to develop his later explanations regarding the behavior of the agents within the system.

Overall Pattern of Responses. Overall, our analysis reveals that students’ registrations were mainly premised on the observations that butterflies were changing color, and that the number of butterflies were increasing overall. The *tricking* and *changing color* registrations were found in 2 of the 10 the students’ responses. Seven students also explicitly mentioned the *fluctuation* registration, i.e., they noticed that the population of the two species were fluctuating. Additionally, when asked to explain where the new butterflies were coming from, half (5 of 10) of the students responded that the new butterflies were either arriving from off-screen or emerging from hiding on the sides on the flowers. We termed this the type of response the *hiding* registration, as this type of response is premised on the interpretation that when butterflies disappear, they hide somewhere off-screen, or when new butterflies appear, they come out form hiding under the flowers. We also found evidence of a similar registration – which we termed *leaving* – in the responses of three students. This registration was premised on the interpretation that when the butterflies disappear from view, it indicates (to the learner) that they travel somewhere outside the on-screen view. The *piling up* registration was invoked by a total of 4 of the 10 students to explain their observations of increases in the population of butterflies of specific colors.

The *hiding, leaving* and *piling up* registrations are evidence of the difficulty students experience in reasoning about population growth and decay using the ideas of reproduction and death. The correct biological mechanism that results in population increase is reproduction, but it is not surprising that the participants would not be able to invoke reproduction as a causal mechanism during these early interactions with the model. This is because, as we reported in the introductory section of the paper, research on early biological knowledge has shown that novice learners at the elementary grades (or below) do not have a lot of experience with the idea or the phenomena of reproduction or death. The number of the students who mentioned each of the seven different registrations is shown below in Figure 4, and Table V below shows examples of sample responses pertaining to each registration.

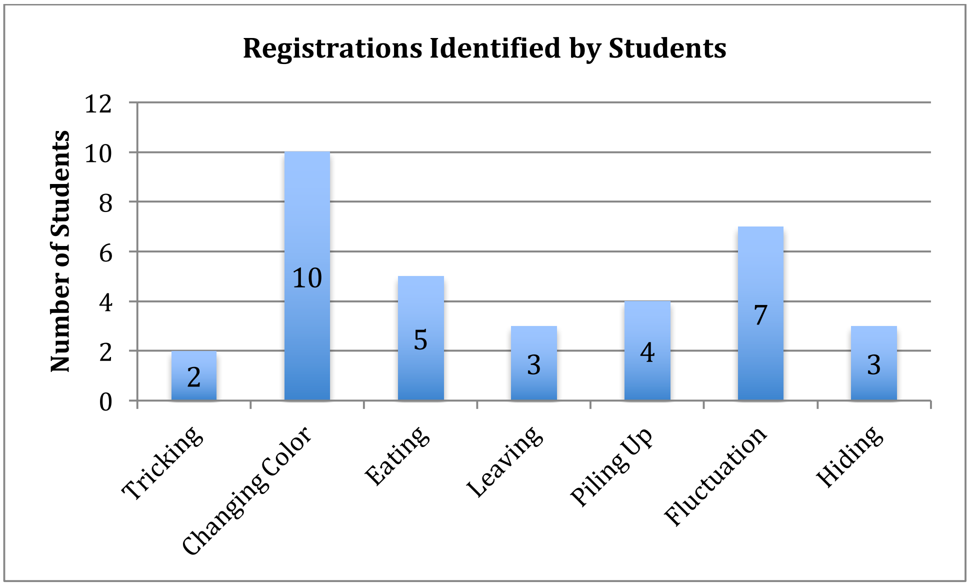


Figure 4: Number of Students' Responses (Y-Axis) Indicating Each Type of Registration

Note that all but one of the registrations identified in the course of Activity 1 were based on agent-level actions or attributes. Only one registration - *fluctuation* - involved observation of the aggregate level effects or outcomes. This is a significant finding, as this shows that although the model displays emergent effects such as population change over generations as well as individual level behaviors, learners primarily tend to focus on the actions of individual agents, unless they are explicitly scaffolded (e.g., by verbal prompting) to focus on aggregate-level effects (such as population change). This supports Levy & Wilensky (2008), who suggested that agent-level reasoning is developmentally prior to aggregate-level reasoning.

Table V: Examples of Different Registrations

Registration	Student Examples
Changing Color	"They (the butterflies) *changing colors*" (Conitra)
Piling Up	"They're (the butterflies) leaving and the butterflies are just *standing in a pile*." (Larry)
Tricking	"They (the butterflies) might change color if the light green one is close by the flower, *it* (the bird) *might think it's* (the butterfly) *the flower if it be still*." (Takeria)
Leaving	"The butterflies *were leaving* and now they're coming back." (Manishe)
Eating	"Because they're (the birds) *eating them* (the butterflies)" (Joseph)
Fluctuation	"It (the butterfly population) *got small then big*." (Olivia)
Hiding	"They were *hiding* under the flowers." (Courtney)

Activity 2: Students' Predictions of Agent Behavior and Aggregate Level Outcomes. During Activity 2, students were asked to make predictions about both the agent-level behaviors and the aggregate-level outcomes they might observe when they ran the model for a second, third or fourth iteration, without changing variables. In general, our analysis reveals that students' responses during this activity were directly based on the registrations they had identified during Activity 1. This is also illustrated below in the cases of Conitra and Larry.

Conitra's Case. In the interview excerpt quoted below, Conitra was asked to make a prediction about what she thinks the agent-level and aggregate-level behaviors might be once she pressed "Setup". This excerpt shows that Conitra's prediction is based on the *changing color* registration that we identified previously during her responses in Activity 1. In lines 10 - 11, she explains that if some of the butterflies did not change color they would probably be easier for the birds to see. Upon further prompting by the interviewer to elaborate her explanation, Conitra explicitly linked *changing color* with *survival* or being *"safe"* (lines 24 and 25), predicting that the butterflies that are capable of camouflaging with the flowers will avoid being eaten and those that are not capable of camouflage will be eaten.

Excerpt 5

Interviewer: Um, based on what you saw when we ran the last program, what do you think is going to happen? What do you think the butterflies' behavior might be?

Conitra: Different.

Interviewer: What do you mean by different?

Conitra: Like, they going to fly in different ways and stuff and some of them might not turn the same color as the flowers.

Interviewer: Okay, why would some of them not turn the same color as the flowers?

Conitra: It would probably be easier for the birds to see them because the birds are the same color.

Interviewer: So you think that the butterflies that are not the same color as the flowers would be easier to see by the birds?

Conitra: [nods head yes]

Interviewer: Okay, um, what might happen to those butterflies that are not the same color as the flowers?

Conitra: They'll get eaten.

Interviewer: They'll get eaten, uh, by who? Who will eat them?

Conitra: The birds.

Interviewer: Okay, the birds will eat them. What about the butterflies that look like the flowers?

Conitra: They won't be able to see them.

Interviewer: Okay, so what will happen, will they be safe or not safe?

Conitra: The ones that are the same color as the flowers will be safe but the ones that are not the same will be not safe.

Larry's Case. Similar to Conitra, Larry's prediction is also based on the "changing color" registration (line 2). In this excerpt, Larry predicted that the butterflies will change to the color of the flowers to blend in with their surroundings (lines 2, 4 & 7). Like Conitra, Larry also stated that the butterflies *want* to change colors - i.e., he also identified *changing color* as an intentional act of the butterfly. However, although he constructed the same causal relationship as Conitra - that camouflage leads to safety - his explanation is comparatively more anthropomorphic than Conitra's. For example, he stated that birds will want to blend in with their surroundings to avoid having rocks thrown at them by children (lines 9 & 10). This is consistent with Larry's earlier observations regarding "tricking" in Activity 1 (e.g., see Excerpt L2, line 10, where Larry stated that the bird might think that the butterflies of the same color are its cousins), suggesting that Larry's explanations are based on both his initial agent-level observations of the model as well as familiar aspects of everyday life.

Excerpt 6

Interviewer: Okay, what do you think is going to happen?
Larry: They'll change colors - like dark green
Interviewer: Why would they change that color?
Larry: Like, they want to change colors to the plants.
Interviewer: Okay, so they want to change to the flower color, why would they want to look like the flowers?
Larry: To blend in.
Interviewer: To blend in, why would they want to blend in?

Larry: So they *won't get hit by a rock that kids try to throw at them.*

Overall Pattern of Responses. A majority of the students' responses to the question "What do you think will happen when you hit 'Go'?" reveal an interesting relationship between registrations (e.g. changing color) and goals or intentions of the relevant agents (e.g., butterflies): that is. the act of changing color is intentional, as the butterflies *want* to change colors to avoid predation by the birds. This relationship between changing color and survival was invoked by 7 of the 10 participants when they were interviewed during Activity 2. For example, when another student, Gerald, was asked to predict what he would see when he ran the model, similar to both Conitra and Larry he responded that "The butterflies are going to change color to not be eaten". Of note is that in Activity 1, when asked to explain what's going on in the model, Gerald simply stated that the butterflies are changing color, thereby indicating the "changing color" registration. Gerald's response in Activity 2; however, indicates that the action indicated by the registration is now associated with a particular goal of the agent: survival. Sample responses of a few students are provided below in Table VI.

Overall, students' responses in Activity 2 had the following characteristics: a) many of them are *need-based*, and b) many of them were based on phenomena that they directly experience on their everyday lives. Both these characteristics are consistent with the literature on early biological knowledge (Inagaki & Hatano, 2002), as well as with the general constructivist approach as suggested by Smith, diSessa & Sherin (1994), which suggests that everyday experiences can provide a productive foundation based on which scientific knowledge can develop.

Table VI: Sample Student Responses in Activity 2

Student	Responses to the Question: What do you think is going to happen when you hit 'Go'?
Takeria	"The butterflies will camouflage so they can hide from the birds"
Gerald	"The butterflies are going to change color to not be eaten"
Courtney	"They will change to green because the flowers are green. They will change because the birds are hungry"
Joseph	"They will change to light or dark blue (the flowers are blue) because if the birds aren't looking, the butterfly might land on a flower and not get eaten."
Camille	"They're (the butterflies) gonna turn green to be the same color as the flowers".

Activity 3: Linking Agent-Level Behaviors to Aggregate-Level Outcomes. In Activity 3 the interviewer first asked the learners to explain the agent-level behaviors they were observing on the screen based on their responses in the previous activities (e.g., why do you think the butterflies changed color?), and then asked the learners to explain aggregate-level effects such as the change in populations of the different species. Overall, learners' responses to these questions reveal that similar to Activity 2, their explanations were based on their initial registrations identified in Activity 1.

Conitra's Case: In the excerpt 7, Conitra correctly identified reproduction as a mechanism for population growth (line 6). She also identified (correctly) that only the successful phenotype - i.e., the butterflies with the same color as the flowers - are the ones that reproduce (lines 5, 6 and 11). Here we see evidence of a causal schema: Camouflage leads to Survival, which in turn leads to Population Increase. Note Conitra's initial registrations - changing color and population increase - are central components of the causal schema.

Excerpt 7

Interviewer: Okay, um, why do you think there are more butterflies now than we started with? You said, remember the butterfly population went down and then it went back up again.

Conitra: Yeah, because they changed the same color as the flower and they coulda had babies.

Interviewer: Okay, so the ones that were the same color as the flowers had babies. And so, the population that we have right now, now if the butterflies that have babies right now kept having babies, what color do you think those babies might be?

Conitra: The same color.

As the interview continues further, Conitra attempts to explain the nature of the dependence between the populations of the birds and the butterflies. In the conversation that ensues (see Excerpt 8), Conitra utilizes her ideas about agent-level behaviors (e.g., eating) in order to explain the aggregate level effects such as change in the populations of the different species, as well as the co-dependence of populations of the birds and the butterflies. For example, when Conitra was asked to explain why the population of the birds was increasing (lines 2 & 3), she explained that the birds were increasing because there were a lot of butterflies (line 4), and that the birds would have a lot to eat (line 7). Here she identifies an agent-level behavior - eating - in order to explain why a higher population of butterflies (prey) can result in a higher population of birds (predator).

Excerpt 8

Conitra: The birds getting higher...

Interviewer: They are getting higher, why do you think the birds are getting higher?

Conitra: There's a lot of butterflies.

Interviewer: Okay, there's a lot of butterflies...Why would that cause the bird population to go up?

Conitra: Because they have a lot to eat.

Interviewer: They...okay, what do you mean by that?

Conitra: They got a lot of butterflies.

Interviewer: They've got a lot of butterflies.

Conitra: They (the birds) keep going down and going up…

Interviewer: Yeah, look at how many, is the bird count getting higher or lower?

Conitra: Higher.

Interviewer: Much higher, yeah. Okay, do you think that, what's going to happen if the birds keep getting higher and higher and higher.

Conitra: The butterflies are going to (inaudible)

Interviewer: The butterflies are going to what?

Conitra: The butterflies are going to get lower, now it's more birds.

Interviewer: Uh huh, There's more birds than butterflies. So you think that the butterfly population is going to get lower. What's happening to it now?

Conitra: It's going down (butterfly population). It's too many birds!

Interviewer: It's too many birds! So what do you think happened?

26 Conitra: They going to eventually become extinct.

27 Interviewer: What's going to become extinct?

28 Conitra: The butterflies.

29 Interviewer: So because they're going to be what?

30 Conitra: Because they out numbered... now the birds are going down.

31 Interviewer: Why are the birds going down?

32 Conitra: Cuz there ain't no more butterflies.

33 Interviewer: Right, because there’s no more butterflies.

As the interview progresses, Conitra further elaborates her explanation of the dependence between the two populations. In line 11, she observes that the population of the birds is fluctuating. As the model runs further, she notices that the butterfly population is decreasing (line 20), and upon prompting to explain what happened, she stated that there are too many birds (line 24). She then predicts that the butterflies are going to be extinct, because they are outnumbered (lines 26 and 30). Finally, she observes that the number of birds is also going down (line 30), and upon prompting, explains that the birds are decreasing in number because of there are no butterflies available for them to eat (line 32).

Throughout this interview, Conitra uses multi-level observations and explanations, i.e., explanations that involve both agent-level behaviors (e.g., eating), as well as aggregate level observations (e.g., change in populations of species) in order to identify and explain a complex dependence of the populations of birds and butterflies on one another. Note that the nature of this dependence, although explained in a qualitative manner by Conitra, is similar to the dependence between predator and prey populations as expressed by the Lotka-Volterra equation, as shown in Figure 5.

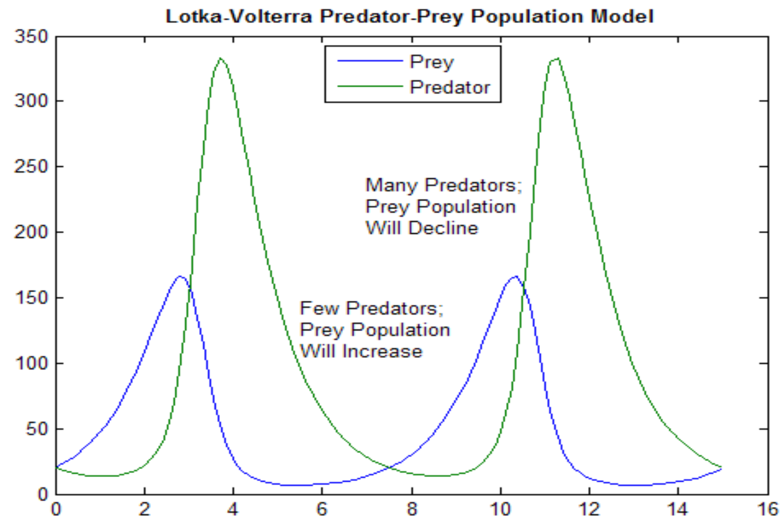


Figure 5: A Prototypical Graph as calculated based on the Lotka-Volterra Differential Equation. The Blue curve represents prey population and green curve represents the predator populations. Y-axis represents population of each species, and X-axis represents time (arbitrary units).

Larry's Case. Similar to Conitra, Larry also demonstrated that he understood that the populations of birds and butterflies were co-dependent, as the following excerpt (9) shows. This excerpt begins with the interviewer asking Larry to explain his observation that there are more butterflies than birds (at a certain moment during the run). In lines 2, Larry explains his observation by stating that the population of the butterflies increases because even though the butterfly population went down initially, the butterflies came back and sustained and increased the population. When the interviewer probed further for an explicit mechanism based on which the butterflies increased their population, Larry stated that the population of the birds and butterflies are inversely proportional (line 4). Larry then re-states the proportionality in lines 6 and 8, as the interviewer asked for clarification in lines 4 and 7, because it was a bit difficult for the interviewer to hear Larry's statement in the line above due to ambient noise.

Excerpt 9

1	Interviewer:	Why are there more butterflies than birds?
2	Larry:	Because the butterflies come back and, like, sustain a population
3		and increase it.
4	Interviewer:	How would the butterflies increase their population?
5	Larry:	If there are less birds, butterflies are increasing but if there are
6		more butterflies then the birds are increasing
7	Interviewer:	Okay so if there are less birds, what's increasing? *[Clarification requested because it was difficult for the interviewer to hear Larry's statement in the line above due to ambient noise]*
8	Larry:	The butterflies.
9	Interviewer:	If there are more birds what would happen to the butterflies?
10	Larry:	They'd be decreasing.
11	Interviewer:	What's happening… [pointing to the butterflies on the screen] it
12		went down, and now what's happening? Which butterflies do we
13		have left?
14	Larry:	Green
15	Interviewer:	Why do we have green ones left?
16	Larry:	Because they're trying to blend in with the flowers.

Note that both the explanations offered by Larry in lines 2, 3, 5 and 6 in Excerpt 9 involve aggregate-level entities – i.e., the population of each species. However, as the interview progressed further, Larry identifies an agent-level interaction that can explain the aggregate-level effect that he observed. In lines 12 and 13, the interviewer asked Larry to explain why butterflies of a certain color (green) were the only butterflies still alive. Larry's explanation (see lines 14 and 16) was based on the *camouflage* registration, as he stated that the green butterflies survived because they blended in with the flowers. Larry's response thus illustrates the causal schema Camouflage → Survival, leading to an increase in species population.

The interviewer then asked Larry to explain the mechanism for butterfly population growth. Larry identified reproduction as the mechanism of population growth (line 4, in Excerpt 10) and upon further prompting, correctly identified that only the successful phenotype - the butterflies the same color as flowers - would be capable of reproducing due to camouflage.

Excerpt 10

Interviewer: Okay, so they're trying to blend in...um, what also happened? We got new butterflies - remember it went down then up. Where did those new butterflies come from?
Larry: Their mothers.
Interviewer: Which butterflies had babies?
Larry: The green ones.
Interviewer: Why did the green ones have babies?
Larry: Cuz, uh, their mothers were already green.
Interviewer: Why were the green butterflies able to have babies? Why are there not any red butterflies having red babies?
Larry: Because there's not any red plants.
Interviewer: Right, so if there were red flowers what kind of butterflies do you think there would be?
Larry: The butterflies would be red.

Overall Pattern of Student Responses. Similar to the Conitra and Larry, all the other students' responses during Activity 3 provided evidence of the causal schema Camouflage leads to Survival. We found that all students used this explanation in order to explain the phenomenon of different survival rates among different species they were observing on the screen. We found two complementary forms of this explanation. Both these forms of explanations were based on the idea that survival (or lack thereof), resulting from camouflage leads to a population increase (or decrease). We identified these causal schemas as *Camouflage →Survival*, which was used to explain the aggregate-level outcome of increase in population of a species, and *No Camouflage → No Survival*, which was used to explain the aggregate-level outcome of decrease in population of a species. Sample responses in which students offer population level explanations related to the causal schema Camouflage →Survival (leading to population increase) and No Camouflage → No Survival (leading to population decrease) are depicted below in Table VII.

Overall, our analysis of students' responses in Activity 3 suggests that in the course of this activity, students were developing explanations of population dynamics by paralleling the actions of the individual agents. Paralleling, as discussed by Levy and Wilensky (2008), is a way for reasoning about complex systems in which students mentally simulate multiple agents acting and interacting concurrently. Paralleling thus involves first recognizing that individual agent-level behaviors are dependent on varying local environmental variables as experienced by the individual agents; thereafter the learner then concurrently applies these agent-level behaviors in order to explain the generation of aggregate-level patterns. For example, student responses in Activity three, as indicated in both Conitra and Larry's responses, reveal that survival of the butterflies is dependent on a particular aspect of the environment – the flower color. Larger aggregate patterns, such as increase in the population of different species of butterflies result due to concurrent changes in the color of individual butterflies that help them camouflage, thereby leading to their increased chance of survival. This indicates that in this step of the learning process, the learners begin to develop multi-level explanations, i.e., they are able to identify specific agent-level action(s) of the butterflies, which can explain relevant aggregate level effects such as of increase in population.

Table VII: Student Responses that demonstrate the causal schemas Camouflage → Survival (leading to Population Increase) and No Camouflage →No Survival (leading to Population Decrease)

Student	Population Increase	Population Decrease
Camille	"The population will go down and then back up because some (the butterflies that are unlike the flowers) will be eaten."	"The butterflies that are not able to blend in will be eaten and then it'll [the number of butterflies] go down."
Joseph	"The population will come back up, the orange ones (flower color is orange) will have babies."	"The population will go down since those butterflies (butterflies that do not match the flowers) have no more camouflage."
Manishe	"The butterflies will turn green (flower color is green) and the population will go up."	*Manishe*:"The population will also go down because some of the birds ate most of the butterflies." *Interviewer*: "Which butterflies did the birds eat?" *Manishe*: "The ones that didn't look like the flowers."
Olivia	"It (the population) will come back up, the pink ones (flower color is pink) will have babies because they won't be dead."	"The population will go down because they (the birds) will eat the butterflies that don't look like the flowers."

Qualitative Differences Between Pre- and Post-Test Responses

Multi-level explanations. Questions 1 and 2 in the pre- and post-test prompted students to reason at the individual level, i.e., about the behavior of individual actors or agents. Question 3 on the other hand, required students to reason about population dynamics, i.e., the aggregate behavior at the level of the species. Our analysis shows that in the pre-test, students were able to reason successfully about the first two questions, whereas many of them found Question 3 to be comparatively more challenging. In their responses to the first question, 7 of the 10 students in the pre-test, and 8 of the 10 students in the post-test provided correct answers. In their responses to the second question, 6 of the 10 students in the pre-test, and 9 of the 10 students in the post-test provided correct answers. In contrast, only 4 of the 10 students provided correct explanations for Question 3 in the pre-test, whereas 9 of the 10 students were able to provide correct explanation in the post-test.

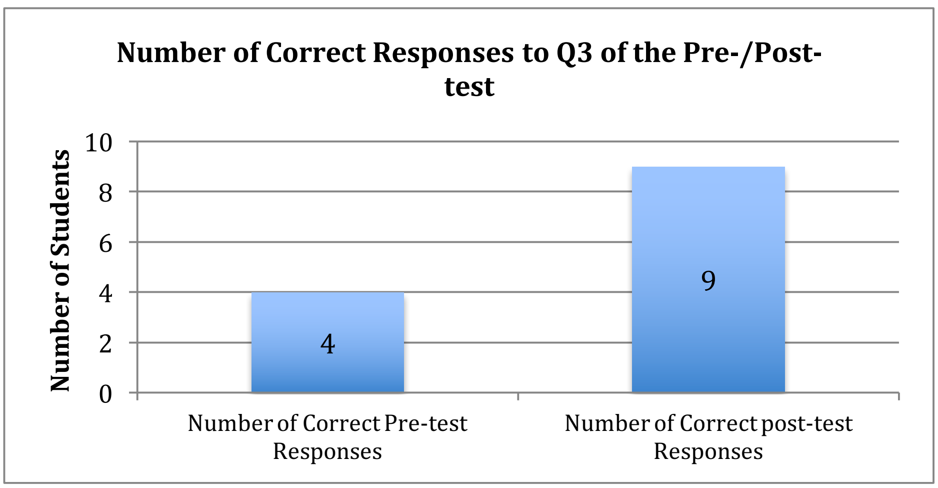


Figure 6: Pre-post comparison of correct responses to Q3 of the pre/post-test.

Student responses to Q3 of the pre-/post-test revealed interesting differences in terms of the use of levels-based perspectives. Prior to their interactions with the model, only 4 of the responses to Q3 in the pre-test provided evidence of agent-aggregate complementarity (for example, see R3 below), whereas the remaining six responses were either based on the agent perspective (see R1 below), or the aggregate perspective (see R2 below).

R1: “The light colored moth would *get caught*.”

R2: “The *population will get big*, big.”

R3: “The light *moths will get eaten* because *they can be seen by the birds*”

Additionally, a large number (4 of 10) of the pre-test responses were limited to stating only what would happen to the population (i.e., the population would either go up or go down), without offering any explanation of the underlying mechanism for the change. In contrast, in the post test, 9 of the 10 responses to Q3 provided correct mechanistic explanations of the population level behavior - i.e., the light moth population would suffer from an absence of lichen and therefore go down, and that the dark moth populations would either benefit or have a neutral effect from an absence of lichen. This indicates that compared to the pre-test, a much greater percentage of students explicitly connected a higher number of agent-level attributes, behaviors and/or interactions to corresponding emergent behavior(s) in the post-test. Even in the case of Manishe, who was able to provide a correct aggregate-level explanation in the pre-test, identified the individual-level attributes of the moths (both species) based on which he explained the change in population size in the post-test. Nearly all students’ post-test responses were indicative of agent-aggregate complementarity.

Table VIII: Sample Students' Responses in Pre- and Post-tests (Each row corresponds to a different student)

	Pre-Test Response (Question 3)	Post-Test Responses (Question 3)
Devon	"All moths are eaten"	"The population will start going down, the light-colored moths will go down because they are easier to see so they might get eaten." "The dark moths will stay up because they are hard to see."
Gerald	"The Great Tit would die because there's no food and the moths would die because there are no trees"	"The would be less light moths because there's no place for them to rest on the trees and they will be seen" The dark moths would grow because there are more trees for them to rest on."
Camille	"The moth population would change because the Great Tit would be able to see all of the moths"	"All of the light moths would die because there is no more lichen, but if the light moths lost their 'lichen color' they would survive. The dark moths would survive. The population of moths would go down but then the dark moths will have babies and the population will go back up."
Olivia	"The population would get big, big"	"The population would go down because they (light moths) would probably get eaten if the Great Tit can see them. The population would go down because the moths like to rest on the lichen."
Manishe	The light moths will go extinct and the dark moths will stay the same.	"The light moths will go down and be extinct because the birds keep eating them" "The dark moth population will go up because they will have babies because they're not eaten."

In order to gain a deeper understanding of the shift in the students' levels-based reasoning in terms of the agent-aggregate complementarity of each response, we additionally coded each written response to Q3 in the pre- and post-test in terms of the number of *correct* agent-aggregate links (AALs), as shown in Table IX. Across all students, the average number of AALs used in generating responses to Q3 of the pre-test was 0.9. In contrast, this average increased to 3.1 in post-test responses.

Table IX: Table Showing Average Number of Agent-Aggregate Links Per Response (Number within parenthesis indicates standard deviation)

	Pre-Test	Post-Test
Average AALs	0.9 (.94)	3.1 (1.37)

Comparison between high and low performing students' responses. When we grouped the responses based on the students' achievement profile (as determined by their class teacher), we found significant between-group differences in their pre-test responses in terms of agent-aggregate complementarity and the number of correct agent-aggregate links. In the pre-test, 4 of the 5 low performing students' responses and 1 of the 5 high performing student responses to Q3 in the pre-test failed to distinguish between the light and dark moth populations. Responses of this type typically involved aggregate-level observations without any mention of the underlying agent-level attributes or behaviors from which these aggregate-level behaviors result. After interacting with the model, all students – both high and low achieving – were able to provide agent-aggregate complementary explanations (Figure 7), in which they identified the population level effect as well as relevant agent-level attributes and/or responsible behaviors.

Our analysis also reveals that in the pre-test, the average number of AALs per response in the case of high-performing students was 1.6 (S.D = 0.55), while the average number of AALs per response in the case of low-performing students was 0.4 (S.D = 0.89). In the post-test, both groups showed a remarkable increase – in the case of high performing students, the mean number of AALs per response increased to 3.2 (S.D = 1.92), while in the case of low performing students, the number increased to 3.0 (S.D = 0.7). This is shown in table X below.

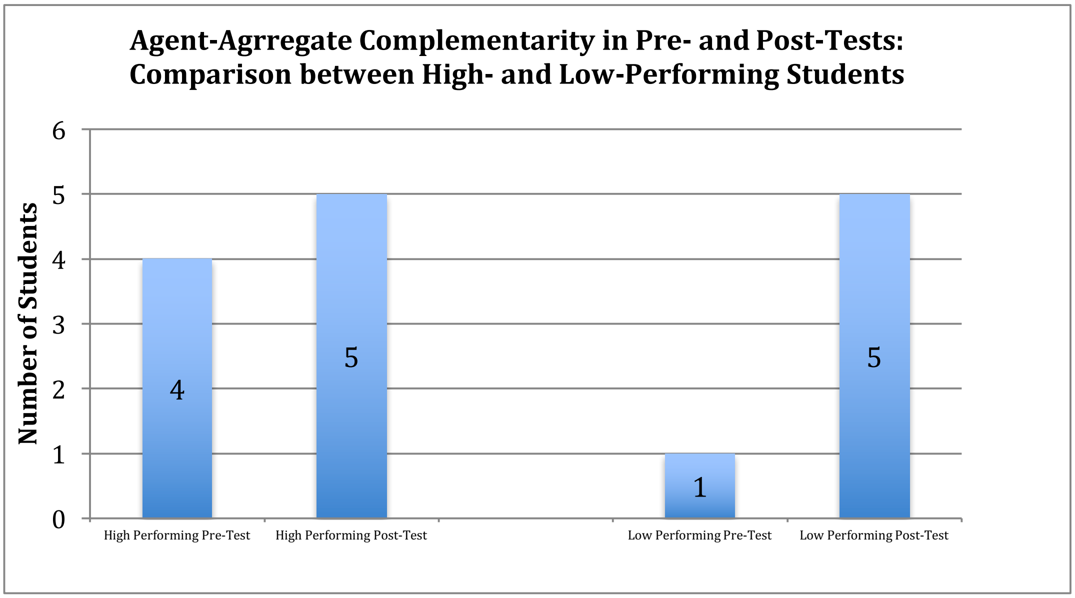


Figure 7: Agent-Aggregate Complementarity in Students' Responses to Q3 Grouped by Their Achievement Profile

Table X: Table Showing Average Number of Agent-Aggregate Links Per Response (Number within parenthesis indicates standard deviation)

	Hi-Performing Students	Low-Performing Students
Pre-test	1.6 (0.55)	0.4 (0.89)
Post-Test	3.2 (1.92)	3.0 (0.71)

Discussion

The Nature and Roles of Learners' Initial Interpretations

The learners' initial registrations observed during Activities 1 & 2 can be understood in light of prior research on how young learners reason about biological phenomena, as discussed in the introductory section of this paper. For example, prior research shows that young learners often reason through novel biological phenomena through an anthropomorphic lens. "Eating", “Tricking” and “Piling Up” are all activities that a young child would be familiar with as both important to survival (safety and eating) and as part of embodied common play activities (piling and tricking). As mentioned in the introductory section of this paper, young novices’ biological reasoning often involves ascribing intentionality to the agents. Students in our study also ascribed intentionality to behaviors of the agents depicted in the model. Students’ use of such intentionality, in turn, were directly linked to anthropomorphic assumptions about the agent’s actions: e.g., birds are *hungry* so they *need* to eat butterflies, and butterflies *do not want to be eaten* so they *need* to change color to either trick the bird or blend in with the flowers. In both these examples, it is evident that students reasoned about the behaviors of birds and butterflies by likening them to humans (i.e., anthropomorphism) and ascribing them intentions.

Children’s difficulties with concepts such as death and reproduction are also evident in their responses to Activities 1 and 2. “Leaving” (i.e., butterflies are leaving the screen) and ‘Piling Up” (i.e., butterflies are piling up one another) were commonly used instead of death and reproduction, respectively, to explain why there were fluctuations in the population level. Our hypothesis is that these difficulties can be explained following Gilbert (1982), who reported that young children reason about biological phenomena based on what they can readily experience, and/or, directly observe. For example, reproduction and death were not explicitly displayed in our model - only their effects were visible in the form of aggregate-level behaviors such as population growth and decline, as well as agent-level phenomena such as appearance and disappearance of individual agents. On the other hand, disappearance - e.g., leaving a place to go somewhere else - is a phenomenon more commonly experienced by young children compared than death. Similarly, some other registrations – such as *piling up* and *changing colors* - are both examples of phenomena that were directly observed and interpreted based on their by students in the model.

However, in Activity 3, all the students mentioned death due to predation as the cause for population decrease of butterflies with colors different from that of the flowers. For example, Larry explicitly identified that birds would eat butterflies of a particular color, as a result of which the overall population of butterflies would go down. We believe that this is because his attention was specifically directed by the interviewer to population-level effects - i.e., the interviewer asked him why populations of butterflies of particular types would change). In contrast, when the same student (Larry) used the *leaving* registration, he was focused on individual butterflies disappearing from screen. This indicates that attending to phenomena at different levels may lead to different types of interpretations for learners.

In the particular context of our MABM, we believe that in attending to the aggregate-level picture of population growth of butterflies with colors close to that of the flowers, (as opposed to mainly focusing on the appearance or disappearance of individual butterflies in Activities 1 and 2), students were able to correctly identify and interpret the phenomenon of survival (or extinction) in terms of camouflage. Given that learners as young as elementary students might tend to reason about biological phenomena using direct experiences, this finding is particularly important for such learners, as different levels of the same phenomena, when displayed in MABMs, may make explicit different types of mechanisms to them.

And finally, the origin of learners' multi-level causal schemas identified during Activity 3 can be traced back to their initial registrations identified during Activities 1 and 2. Consider for instance the explanation: *Camouflage leads to Survival*. Here, camouflage can be regarded as a corollary of the registration "changing color" which was elicited by all of the students during their interview responses in Activities 1 and 2. In light of this finding, as well as the previous finding that some of the registrations learners mentioned during Activities 1 and 2 were canonically incorrect, we believe that it is important for educational designers of MABMs to pay particular attention to the learners' initial interpretations of the phenomena depicted in the MABM, and think carefully about how to scaffold the learners in terms of focusing their attention to different levels of the depicted phenomena.

Direct Causal Reasoning in Learners' Explanations

Several scholars have argued that knowledge elements that involve direct causation (e.g., simple causal relationships such as A causes B), or direct schema (e.g., knowledge elements that involve attributes of actions of individual agents) hinder productive learning in the context of learning natural selection and other emergent phenomena. For example, Chi (2005) wrote:

> "This skewed misrepresentation *[direct causal reasoning]* may have an innate source, in that even infants seem to understand a direct kind of causality. This innate predisposition to interpret all processes as a direct kind may be another source for the robustness of misconceptions that makes them difficult to overcome."

Our findings on the other hand, reveal that direct causal relationships between two entities of variables played an important role in learners' explanations, as evident from students' explanations both in the pre- and post-tests, as well as during their interaction with the model. For example, the causal schema *camouflage leads to survival* indicates a direct causal relationship (or a direct schema, see Chi, 2005) between the behavior of an agent, and an aggregate level outcome, and students' anthropomorphic reasoning (e.g., the birds are hungry so they need to eat the butterflies) are also examples of such reasoning. All the registrations noted in this study would be examples of direct schema, as they are primarily comprised on attributes of or actions of individual agents. While not all the examples of direct causal reasoning or direct schema may play productive roles in the process of students' conceptual development (e.g., the *leaving* registration), our study suggests that it would be incorrect to label direct causal reasoning as "skewed misrepresentations" for learning about emergent phenomena, at least in the context of the introductory aspects of emergence in the context of predator-prey dynamics.

Conceptual Growth in Terms of Agent-Aggregate Complementarity

As mentioned in the beginning of the paper, one of the key findings in the literature on complex systems education is that reasoning about complex systems involves being able to reason about multiple levels in which the phenomena take place. All students in this study showed evidence of conceptual growth along this dimension. The comparison between students' pre and post-test responses shows evidence of students' conceptual growth in terms of being able to identify and differentiate between members of species based on salient traits and selection pressure, as well as being able to develop multi-level, agent-aggregate complementary explanations of population change over time. We found that all students were able to provide agent-aggregate complementary explanations of population-level phenomena in the post-test, compared to 5 of the 10 students in the pre-test. We also found that even though high-achieving students provided on average more correct agent-aggregate complementary explanations in the pre-test, after their interactions with the model, the average number of AALs per response was almost identical (almost equal to 3) for students in both groups in the post-test.

Limitations and Future Work

Despite the gains in student thinking, we acknowledge that the small sample size and the pullout nature of this study limit our ability in predicting classroom-wide learning gains. A natural next step for us is therefore to conduct classroom-level studies with larger sample sizes. Such a study would also involve developing curricular materials and teacher guides to scaffold students' interaction with the simulation based on the lessons from our present study. For example, given that our study highlights the importance of the different levels of phenomena that students pay attention to, teachers can scaffold students' interactions with the simulation by providing prompts to focus their (students') attention to specific levels of the represented phenomena, as appropriate to the task at hand. We are currently in the process of developing curricular materials for classroom-wide studies in the near future.

It is also important to note that scientific expertise develops over a span of several years through sustained immersion in the generative and authentic scientific practices (Lehrer & Schauble, 2006). This study is only a small (but important) step in that direction. While we show that 4th graders can indeed develop a deep understanding of at least some introductory aspects of natural selection, it leaves us with the question "what happens next in their learning trajectory". For example, how can elementary students develop multi-level explanations of more complex aspects of evolutionary phenomena, e.g., feedback effects that are also characteristic of predator-prey scenario using MABMs? Answering the above question requires an inherently developmental approach to the design of a MABM-based learning progression, and as such is part of an ongoing research agenda within our lab.

Finally, to summarize, our study highlights the importance of learners' intuitive knowledge and initial interpretations of phenomena represented in the simulation in the process of their knowledge construction – a finding that we believe may have implications for designers of educational simulations of scientific phenomena (in general). Specific to the context of designing educational MABMs, as we have argued earlier, our study suggests that it is important for educational designers of MABMs to pay particular attention to the learners' initial interpretations of the phenomena depicted in the MABM, and think carefully about how to scaffold the learners in terms of focusing their attention to different levels of the depicted phenomena. This is important because the complexity of information embodied in the dynamic simulation displays in MABMs might make it challenging for learners to focus on or identify appropriate elements in the simulation. Indeed, Basu, Biswas & Sengupta (2011) showed that when middle students engage in learning about ecology using MABMs, they often require explicit scaffolding in order to focus their attention to specific levels of the phenomenon appropriate to the task at hand, and in absence of such scaffolding, they are unable to develop agent-aggregate complementarity. The results we have presented here further show that i) children's noticing, observations and inferences during their interactions with MABMs are not independent of their unschooled intuitive knowledge, and ii) elements of their repertoire of intuitive knowledge can indeed be productively bootstrapped for the development of deeper, canonical understanding of the target phenomena. We therefore believe that during the design process, it is important for educational designers to iteratively refine their models and learning activities based on feedback from the learners, and furthermore, during these interactions with the learners, they should pay careful attention to learners' initial knowledge and understandings of the phenomena represented in the models. Our study suggests that doing so can help educators design instructional supports for classroom use of MABMs that in tune with the constructivist perspective outlined by Smith, diSessa & Sherin (1994) and Sengupta & Wilensky (2009), will bootstrap, rather than discard the repertoire of intuitive knowledge that novice learners bring in with them to the instructional setting.

References

Anderson, D., Fisher, K and Norman, G. (2002). Development and evaluation of the conceptual inventory of natural selection. *Journal of Research in Science Teaching*, *39 (10),* 952-978.

Abrahamson, D. (2004). *Keeping meaning in proportion*. Unpublished doctoral dissertation, Northwestern University, Evanston, IL.

Basu, S., Biswas, G., & Sengupta, P. (2011). Scaffolding to Support Learning of Ecology in Simulation Environments. In: Biswas, G., & Bull, S. (Eds.). *Lecture Notes in Computer Science*, *AIED 2011,* 6738/2011, 417-419.

Bernstein, A.C. (1975). Children's concepts of how people get babies. *Child Development*, *46 (1)*, 77.

Blikstein, P., & Wilensky, U. (2009). An atom is known by the company it keeps: A constructionist learning environment for materials science using multi-agent simulation. *International Journal of Computers for Mathematical Learning*.

Carey, S. (1985). *Conceptual Change in Childhood*. Cambridge, MA: Bradford Books, MIT Press.

Chi, M. T. H. (2005). Commonsense Conceptions of Emergent Processes: Why Some Misconceptions Are Robust. *Journal of the Learning Sciences, 14*, 161-199.

Chi, M. T. H. and Ferrari, M. (1998). The nature of naïve explanations in natural selection. *International Journal of Science Education*. *20 (10),*1231-1256.

Coley, J.D. (2000). On the importance of comparative research: The case of folkbiology. *Child Development*, *71*, 82–90

Darwin, C. (1989). *On the origin of species.* New York: Gramercy (Republication of the first edition, published in 1859).

Denzin, N. K., & Lincoln, Y. S. (2000). *Introduction: The discipline and practice of qualitative research.* In N. K. Denzin & Y. S. Lincoln (Eds.), Handbook of qualitative research (2nd ed., pp. 1-28). Thousand Oaks, CA: Sage.

diSessa, A. (2000). *Changing Minds: Computers, Learning, & Literacy*. MIT Press, Cambridge, MA.

Dodick, J. & Orion, N. (2003). Introducing evolution to non-biology majors' via the fossil record. *The American Biology Teacher*: *65 (3)*, 185-190.

Forbus, K., & Gentner, D. (1986). *Learning physical domains: Towards a theoretical framework.* In R.S. Michalski, J.G. Carbonell, & T.M. Mitchell (Eds.), Machine learning: An artificial intelligence approach (Vol. 2, pp. 311-348). Los Altos, CA. Morgan Kaufmann.

Gould, S. J. (1996). *Full House: The spread of excellence from Plato to Darwin.* New York: Harmony Books.

Gilbert, J.K. (1982) Children's Science and its Implications for Teaching. *Science Education*, *66 (4)*, 623.

Gomm, R., Hammersley, M. and Foster, P. (2000). *Case Study Method: Key Issues, Key Texts.* Sage, Thousand Oaks, CA.

Hammer, D. (1996). Misconceptions or p-prims: How may alternative perspectives of cognitive structure influence instructional perceptions and intentions? *Journal of the Learning Sciences,* 5 (2), 97-127.

Hendrix, J.R., Mertens, T.R., & Baumgartner, R.S. (1981) Individualizing Instruction through Concept Assessment. *American Biology Teaching*, *43(5),* 246-253.

Hmelo-Silver, C., Marathe, S., & Liu, L. (2007). Fish swim, rocks sit, and lungs breathe: Expert-novice understanding of complex systems. *Journal of the Learning Sciences, 16*(3), 307-331.

Holland, J. (1998). *Emergence: from chaos to order.* New York: Addison-Wesley Longman.

Ingagaki, K. & Hatano, G. (2002). *Young Children's Naïve Thinking About the Biological World.* New York: Psychology Press.

Jacobson, M., & Wilensky, U. (2006). Complex systems in education: Scientific and educational importance and implications for the learning sciences. *The Journal of the Learning Sciences, 15 (1),* 11-34.

Keil, F.C. (1992). The origins of autonomous biology. In M.R. Gunnar & M. Maratsos (Eds.), *Modularity and constraints in language and cognition. The Minnesota Symposia on Child Psychology* (Vol. 25, pp. 103-137). Hilldale, NJ: Erlbaum.

Kelemen, D., (1999). *The origins of teleological thought.* In: Corballis, M., Lea, S., (Eds.), Descent of Mind. Oxford University Press, Oxford.

Klopfer, E., Yoon, S., & Perry, J. (2005). Using palm technology in Participatory Simulations of complex systems: A new take on ubiquitous and accessible mobile computing. *Journal of Science Education and Technology, 14(3)*, 285-298.

Klopfer, E., Yoon, S., & Um, T. (2005). Teaching complex dynamic systems to young students with StarLogo. *Journal of Computers in Mathematics and Science Teaching, 24(2),* 157-178, 2005.

Lehrer, R., & Schauble, L. (2006). Scientific thinking and science literacy. In W. Damon, R. Lerner, K.A. Renninger, and I. E. Sigel, Eds. *Handbook of child psychology, 6th Edition, Volume 4: Child psychology in practice* (pp. 153-196). Hoboken, NJ: John Wiley and Sons.

Levy, S. T., & Wilensky, U. (2008). Inventing a "mid-level" to make ends meet: Reasoning through the levels of complexity. *Cognition and Instruction, 26(1)*, 1-47.

Maxwell, J.C. (1871 (2001)). *Theory of Heat.* New York: Dover.

Mitchell, M. (2009). *Complexity: A Guided Tour*. Oxford University Press, New York, NY.

Papert, S. (1980). *Mindstorms: Children, computers, and powerful ideas*. Basic Books, New York, NY.

Penner, D. E. (2000). Explaining systems: Investigating middle school students' understanding of emergent phenomena. *Journal of Research in Science Teaching*, 37, 784–806.

Petri, J. & Niedderer, H. 1998 A learning pathway in high-school level quantum atomic physics *Science Education,* 20, 1075–88

Piaget, J., 1929. The child's conception of the world. Routledge and Kegan Paul, London.Reiner, M., Slotta, J.D., Chi, M.T.H. and Resnick, L. (2000). Naïve physics reasoning: a commitment to substance-based conceptions. *Cognition and Instruction*, 18 (1), 3-34.

Resnick, M. (1994). *Turtles, termites and traffic jams: Explorations in massively parallel microworlds.* Cambridge: MIT Press.

Roschelle, J. (1991). *Students' construction of qualitative physics knowledge: Learning about velocity and acceleration in a computer microworld.* Unpublished doctoral dissertation, University of California, Berkeley.

Sengupta, P., & Wilensky, U. (2009). Learning electricity with NIELS: Thinking with electrons and thinking in levels. *International Journal of Computers for Mathematical Learning, 14*(1), 21-50.

Sengupta, P., & Wilensky, U. (2011). Multi-Agent-Based Modeling and Learning Electricity: Design and Epistemological Issues. In Khine, M.S., & Saleh, I.M (Eds.). *Dynamic Modeling: Cognitive Tool for Scientific Enquiry*. Springer, New York, NY.

Simons, H. (Ed) (1980). *Towards a Science of the Singular: Essays about Case Study in Educational Research and Evaluation.* CARE Occasional publications, No. 10, Norwich, Centre for Applied Research in Education, University of East Anglia.

Smith, J., diSessa, A., & Roschelle, J. (1993/1994). Misconceptions reconceived: A constructivist analysis of knowledge in transition. *The Journal of the Learning Sciences*, *3 (2),* 115-163.

Solomon, G., Johnson, S., Zaitchick, X. & Carey, S. 1996. Like father, like son: young children's understanding of how and why offspring resemble their parents. *Child Development*,*67*, 151-171.

Strogatz, S. (2003). *Sync: The emerging science of spontaneous order.* Hyperion Books. New York, NY.

Taber, K. S. (2008). Exploring Conceptual Integration in Student Thinking: Evidence from a case study. *International Journal of Science Education*, *30 (14)*, 1915 – 1943

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

Wilensky, U. (2003). Statistical mechanics for secondary school: The GasLab modeling toolkit. *International Journal of Computers for Mathematical Learning*, *8(1),* 1-41 (special issue on agent-based modeling).

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.

Wilensky, U., & Resnick, M. (1999). Thinking in levels: A dynamic systems perspective to making sense of the world. *Journal of Science Education and Technology, 8(1).*

Wilensky, U. & Stroup, W. (1999). HubNet. [Computer Software]. Evanston, IL. The CCL, Northwestern University. http://ccl.northwestern.edu/netlogo/hubnet.html

White, P., A. (1997). Naïve ecology: Causal judgments about a simple ecosystem. *The British Journal of Psychology*, *88*, pp.219-233

Yin, R.K. (1994). *Case study research: Design and methods (2nd ed.).* Applied Social Research Methods Series, 5. Thousand Oaks: Sage Publications.

CHAPTER II

DEVELOPMENT OF MECHANISTIC REASONING AND MULTI-LEVEL EXPLANATIONS OF ECOLOGY IN 3RD GRADE USING AGENT-BASED MODELS[2]

Introduction

Ecological systems are examples of complex systems, and emergence is a key characteristic of such systems (Holland, 1999; Mitchell, 2009; Chi, 2005; Wilensky & Reisman, 2006; Danish et al., 2011). By emergence we mean aggregations of simple, local interactions between many individual actors which give rise to complex and often counterintuitive global patterns. Researchers have shown that students at all levels find understanding emergent processes challenging (Chi, 2005; Resnick & Wilensky, 1998). Agent-based computational models (ABMs) have been shown to be successful in helping novices understand complex ecological systems (Resnick, 1994; Wilensky & Resnick, 1999; Klopfer, Yoon & Perry, 2005; Klopfer, Yoon & Um, 2005; Danish, 2014). The term "agent" in the context of ABMs indicates individual computational objects or actors. It is the behaviors and interactions between these agents as well as elements of the environments in which these agents are situated, that give rise to emergent, system-level behavior (*e.g.*, the formation and movement of a traffic jam or the spread of disease). Each agent in an ABM makes it own decision. This implies that the overall, emergent patterns represented in the simulations do not result from averaging over a population; instead, they result from the aggregation of the outcomes of individual-level decisions of multiple agents. Studies show that curricula which utilize ABMs can help students understand

[2] This chapter was published in *Science Education* in 2016

complex systems and emergence by grounding emergent phenomena in terms of their embodied, agent-level intuitions (Resnick, 1994; Wilensky & Resnick, 1999; Klopfer, Yoon & Um, 2005; Danish, 2014).

In science classrooms, students either interact with pre-developed agent-based computational models (e.g., Danish, 2014), or they program and build agent-based models from scratch (e.g., Wilensky & Reisman, 2006). While the later approach can provide a deeper pathway into computational modeling (e.g., Sengupta et al., 2015), even in cases where students interact with pre-developed agent-based models (e.g., Danish, 2014), they can still engage in deep investigations of the relevant aspects of complexity by engaging in *multi-level,* model-based explorations of, and reasoning about the relevant phenomena. Students typically accomplish this by a) designing and conducting simulation experiments by selecting and testing agent-level and environmental variables, and b) interpreting aggregate-level outcomes in terms of the relevant agent-level interactions (Wilkerson-Jerde & Wilensky, 2014; Danish, 2014; Dickes & Sengupta, 2013; Goldstone & Wilensky, 2008; Levy & Wilensky, 2008).

It is, however, only recently that researchers have focused on integrating ABMs in elementary classrooms with the specific goal to teach students about ecological systems (Danish, 2014; Peppler et al., 2010). Danish and colleagues argued that educational designers should shift their research focus from the developmental constraints of young children to the design of activity systems in which young children can productively engage. That is, rather than investigating children's readiness to learn using ABMs (or conversely, investigating whether or not ABMs are appropriate for children), Danish and colleagues argue that we should be asking how can we design ABM-supported curricular activities for children to investigate complex phenomena. In line with this argument, we believe that there is sufficient evidence that elementary-aged children can both successfully engage in scientific inquiry using ABMs and develop a deep understanding of complex biological phenomena (Danish et al., 2011; Danish, 2014; Dickes & Sengupta, 2013). Furthermore, we believe that this body of work provides useful guidelines for the design of new learning environments that integrate ABMs in elementary education. We discuss these guidelines later, and use them as pedagogical foundations for our work.

In this paper, from the design perspective, we build on Danish and colleagues' work with kindergarten through second grade students by further problematizing the issue of designing and implementing a new learning environment for 3rd grade students in the domain of ecology. Specifically, we seek to identify how embodied modeling (Phase 1), the generation of mathematical inscriptions (Phase 2), and inquiry using ABMs (Phase 3) can be integrated in a learning environment to support 3rd grade students' development of progressively more sophisticated explanations of interactions and inter-relationships within an ecosystem. The ecosystem we focus on bears some similarities with Danish and colleagues' studies: they both involve foraging. However, we focus on different system-level outcomes. While Danish's work focuses on communication between bees, our work emphasizes the following curricular foci in a butterfly-flower-bird ecosystem: foraging, predation, camouflage, population survival, change in energy, and the generation of mathematical representations. As we explain later, these curricular foci were chosen based on the relevant curricular science standards through our partnership with the participating teacher.

Along the dimension of analysis of student learning, we make the following contributions. First, while several frameworks for analyzing student reasoning of complexity have been proposed (Wilensky & Reisman, 2006; Stroup & Wilensky, 2014; Hmelo-Silver & Pfeffer, 2004; Chi, 2005; Wilkerson-Jerde & Wilensky, 2014), we present a more nuanced analysis of the progressive refinement of children's explanations about ecosystems using the lens of mechanistic reasoning (Machamer, Darden, & Carver, 2000; Russ, Scherr, Hammer, & Mikeska, 2008). We adopt the lens of mechanistic reasoning (Russ et al., 2008), which proposes several categories for analyzing causal explanations to identify the different forms of sophistication in their explanations about the ecosystem, particularly those concerning the attributes, behaviors and interactions of individual agents. Although researchers have begun thinking about analyzing students' mechanistic explanations in the context of learning about ecology (Danish et al., 2011; Hmelo-Silver & Eberbach, 2014), we specifically propose that Russ et al. (2008)'s framework can be adopted to conduct a nuanced analysis of student learning in this domain, particularly when children use agent-based models. Second, we highlight the importance of embodied modeling in the learning trajectory of students. Although we do not present a controlled study that compares students participating in embodied modeling activities with students in contrasting conditions (such as only-simulation, or non-embodied-activity plus simulation), we do present a contrasting case which highlights the nuances and importance of embodied reasoning. Third, we also present an analysis of student reasoning and discourse about forms of complexity that Simon (1977) termed *loosely coupled events*, which require reasoning beyond a single causal mechanism. Finally, we show that after interacting with our learning environment, students were better able to identify interrelationships pertaining to producer-consumer energy flow in food webs in the post-assessment compared to the pre-assessment.

Theoretical Background

Modeling as Learning Science

Conceptual development in science is inseparably intertwined with the development of epistemic and representational practices such as modeling (Giere, 1988; Latour, 1999, Lehrer &

Schauble, 2000, 2006; Nersessian, 1992). Considered the *key* epistemic and representational practice of the sciences (Duschl et al, 2007; Nersessian, 1992 & 1998; NRC, 2008), the practice of modeling involves the generation of inscriptions that highlight aspects of a scientific phenomenon (Lehrer & Schauble, 2006a; Rapp & Sengupta, 2012). Researchers have argued that supporting the development of modeling practices in young learners requires designed learning environments where learners have opportunities to interact with representational forms that offer increasingly more sophisticated explanations of the phenomena (Lehrer & Schauble, 2010, 2011; Enyedy, 2005). Researchers have also shown that as children engage in iterative modeling activities, the process of progressively refining their representations of some aspect of the world can contribute to a deeper understanding of a domain (Gravemeijer, Cobb, Bowers, & Whitenack, 2000; Stevens & Hall, 1998; Lehrer & Pritchard, 2002; Lehrer & Schauble, 2006b). Termed the science-as-practice perspective (Lehrer & Schauble, 2006a; NRC, 2008), this body of research suggests that science educators should focus not only on supporting the development of target conceptual ideas, such as structure-function relationships in an ecosystem, but also on supporting the development of modeling and inscriptions as the practice through which these ideas may develop.

Our study is grounded in this perspective. Our learning activities are sequenced such that the embodied and teacher-led modeling and representational activities scaffold the more complex forms of representations students encountered during their investigations with the ABMs. However, it is important to note that, in our study, the variability in the forms of representations that students generated was limited due to the nature of the instructional approach adopted by the classroom teacher, and we discuss this issue further in the section on researcher-teacher partnership. Nonetheless, in our case, student-generated representations during Phases 1 and 2 provide a bridge between representational forms that they have used prior to the study (bar graphs), albeit to represent linear systems such as motion, and those they used in the simulated models of complex systems explored in Phase 3.

Embodied Cognition & Its Role in Learning and Modeling Systems Biology

Recent studies of students (Danish, 2014; Wilensky & Reisman, 2006) as well as scientists in action (Chandrasekharan, 2009; Chandrasekharan & Nersessian, 2014) have highlighted the importance of embodied cognition and simulations in modeling systems biology. As Alibali & Nathan (2012) have pointed out, even though there is not yet a unified theory of embodiment, scholars of embodied cognition generally agree that mental processes are mediated by body-based systems, including body shape, movement, and scale; motor systems, including the neural systems engaged in action planning; and the systems involved in sensation and perception (Dreyfus, 1996; Glenberg, 2010).

Relevant to our paper, three forms of embodied cognition have been shown to be productive in learning and modeling systems biology: Incorporation, Resonance and Egocentrism. In *incorporation,* one's body schema is extended to include external components, such as tools, models and experimental equipment (Ingold, 2013; Chandrasekharan & Nersessian, 2014). In *resonance,* the perceived and imagined dynamics of external systems are replicated using the motor system (Chandrasekharan, 2009). Both of these forms of embodiment involve the coupling between internal imagination (e.g., embodied knowledge and mental models) and external representations, and are often useful in scientific work, because they allow for the testing of imagined 'what-if' scenarios involving many variables, and the discovery of new variables not previously imagined (Chandrasekharan & Nersessian, 2014). Although students in our study did generate external representations during modeling, the learning environment and the curricular design employed in our study emphasizes the third form of embodiment: *egocentric*. In this form of embodiment, one's body is imagined as a biological agent's body. The works of Wilensky & Reisman (2006), Danish (2014), Wagh & Wilensky (2014), and Dickes and Sengupta (2013) fall in this category. In these studies and in ours, learners use agent-based models to investigate emergent outcomes in ecosystems by imagining themselves as biological agents. That is, they investigate and develop explanations of system-level, emergent behaviors from the perspective of agents within the system. A key argument supported by these studies is that thinking like the agent provides learners an intuitive pathway in exploring emergent outcomes of the system. Evelyn Fox Keller's biography of the biologist Barbara McClintock supports this claim, citing evidence that thinking like the agent (*e.g.*, a chromosome) enabled McClintock to make significant advances in her research on human genetic structures (Keller, 1983).

Forms of Complexity Explored in Our Study

Survival and Population Growth. The change of populations of different species and their interdependence have been identified as *emergent* patterns in ecosystems (Wilensky & Reisman, 2006; Wilkerson-Jerde & Wilensky, 2014; Dickes & Sengupta, 2013). The study of population growth and decay involves reasoning about the behaviors that cause populations to fluctuate, such as birth, death, immigration, competition for resources, or population density (Wilkerson-Jerde & Wilensky, 2014; NRC, 2008; Sandholm, 2010). Typically, mathematical models of population growth are part of most high school curricula (Wilkerson-Jerde & Wilensky, 2014; American Association for the Advancement of Science, 1993; Common Core State Standards Initiative, 2010; NRC, 2012), however, we believe that it is possible for younger children (e.g., 3rd and 4th graders) to begin to reason productively about at least some aspects of population growth and decay. Our previous work has shown that learners as young as 4th grade can, when appropriately scaffolded, reason about factors that affect changes in populations (Dickes & Sengupta, 2013). In this paper, we continue this work by focusing on even younger children (3rd graders).

Loose Coupling. Simon (1977), who first proposed the levels-based perspective for analyzing complex systems, argued that complex systems exhibit two forms of coupling. There is “vertical coupling" between levels, in that higher levels are composed of the lower levels. There is also “horizontal coupling" within levels in the form of communication and interaction between subsystems of the same hierarchic level (Simon, 1977; Mitchell, 2009). “Loose coupling” is a form of horizontal coupling in which each component of the horizontally coupled sub-systems operates independently of the internal mechanisms of others. That is, in such interactions, only the input(s) each sub-system requires, and the output(s) it produces are relevant for the emergent, system-level behavior. This in turn implies that the same outputs can be obtained from the same inputs by two or more different paths, a characteristic that Simon (1977) termed 'functional equivalence'.

Functional equivalence is a well-studied phenomenon in the domain of ecology (Hubbell, 2001; Hubbell, 2005; Zamora, 2000). It is defined as the existence of multiple, sufficient explanations for the same community-level ecological phenomenon (Chave, Muler-Landau & Levin, 2002; Purves & Pacala 2005). A cornerstone of “neutral theory” (Hubbell, 2005; Zamora, 2000), it is based on the observation that in many cases, different species within an ecosystem have no niche differences and thus multiple organisms can account for the same community-level effects. Functional equivalence is evident in our study. In our embodied modeling activity there are two species of butterflies: one with a long proboscis, and one with a short proboscis;

and two species of flowers: one with a long nectar sac, and one with a short nectar sac. Within this system, different forms of functional equivalence emerge as butterflies forage for nectar. One form of functional equivalence occurs within the short flower-butterfly system. In this subsystem, the butterflies, although structurally different, are functionally equivalent in terms of being able to obtain nectar from the short flower.

Food Chains. In the context of ecosystems, energy flow and matter cycling, often represented in the form of food chains (see Figure 11 for an example), have been shown to be challenging to understand for K12 students. Researchers have argued that this is due to the difficulties students face in relating the multiple levels of the phenomena (organism, cellular and molecular), in absence of suitable instruction (Lin & Hu, 2003; Eilam, 2012; Brown and Schwartz, 2009). For example, researchers have argued that due to lack of knowledge about underlying mechanisms at a micro level, many students do not understand that the prerequisite of plant placement as first in webs is due to photosynthesis, or that a chains' order is due to evolutionary processes and environmental conditions (Alparslan, Tekkaya & Geban, 2003; Liu and Lesniak, 2006). Reiner and Eilam (2001) have also pointed out that students may also erroneously use egocentric reasoning, and place humans at the top of feeding chains (Dagher & BouJaoude, 1997). Based on these studies, we therefore believe that being able to identify individual level behaviors and interactions in an ecosystem may improve students' understanding of food chains in similar ecosystems. Although our curricular activities do not focus on teaching children to represent food chains and webs, our pre- and post-assessments did ask children to create such representations.

Guidelines for Designing ABM-based Learning Environments in Ecology

Studies show that when learners use agent-based models to investigate complex phenomena in physics and biology, they can develop an understanding of the relationships between the agent-level behaviors and the aggregate-level outcomes of the system (Blikstein & Wilensky, 2009; Klopfer & Resnick, 2003; Dickes & Sengupta, 2013; Danish et al., 2011). Prior research has predominantly focused on how students can learn about emergent phenomena in ecology using ABMs in middle school, high school, and beyond (Kuhn & Reiser, 2006; Wilensky & Reisman, 2006; Tan & Biswas, 2007; etc.). The use of ABMs in elementary science curricula is relatively recent. These studies show that: 1) ABMs can help students bootstrap their intuitive knowledge of agent-level behaviors to develop understandings of emergence in ecology (Danish, 2014; Dickes & Sengupta, 2013; Danish et al., 2011), and 2) appropriate instructional support is integral when incorporating ABMs into science curricula (Basu, Sengupta & Biswas, 2015; Danish et al, 2011; Peppler et al, 2010).

In the studies conducted by Danish and his colleagues, students interacted with a range of activities—both computational and non-computational—in order to develop an understanding of the emergent nature of collection of nectar by honeybees. These activities included individual drawings, creation of skits, engaging in the BeeSim participatory simulation, and playing a custom board game designed specifically for their study (Danish et al., 2011; Danish, 2014). These activities were designed to help students engage with different aspects of the honeybee system. For example, students' drawings were expected to help them think about the bee's anatomic structure, and the participatory skits and simulation were intended to help them think about the inherent challenges in searching for nectar and the benefit of the bee dance in simplifying this search. The agent-based simulation (BeeSign) supported students' engagement with the hive behaviors at the aggregate level through a focus on emergence (Danish et al., 2011). Furthermore, students' explanations about structure, behavior and function of different elements of the honeybee system later in the curricular unit were more refined compared to their earlier explanations (Danish, 2014).

Frameworks for Analyzing Reasoning about Complexity

Levels-Based Reasoning. Researchers have identified two primary levels of reasoning that novices and experts utilize when explaining complex systems: agent-level and aggregate-level reasoning (Abrahamson & Wilensky, 2004; Blikstein & Wilensky, 2009; Wilensky & Resnick, 1999; Sengupta & Wilensky, 2009; Jacobson & Wilensky, 2006). It has been argued that agent-level reasoning develops before aggregate-level reasoning (Goldstone & Wilensky, 2008; Levy & Wilensky, 2008). When students use agent-based-based models to investigate and understand complex phenomena, the instructional goal is to support the development of a particular form of explanation: agent-aggregate complementary explanations (Abrahamson & Wilensky, 2004; Stroup & Wilensky, 2014). These are explanations in which learners explain a pertinent aspect of the emergent phenomenon in terms of relevant individual-level or agent-level attributes and relationships. In the context of population dynamics of ecosystems, Dickes & Sengupta (2013) provided some examples of agent-aggregate complementary explanations, where 4th graders explained change in the population of butterflies and birds (e.g., the number of birds is increasing and the number of butterflies is decreasing) in terms of agent-level behaviors (e.g., birds are eating butterflies), and camouflage (e.g., light colored butterflies are harder to see by birds).

Structure Behavior and Function (SBF) Framework. Hmelo-Silver and colleagues (Hmelo-Silver, Marathe, & Liu, 2007; Hmelo-Silver & Pfeffer, 2004) proposed a framework for investigating expert-novice differences in understanding complex systems based on the structures, behaviors, and functions (SBF) of different elements in the complex system. “Structures” describe the configuration of the components and subcomponents of the system and articulate their connections. “Functions” represent the outputs of those structures or, more specifically, the purpose of an element within the system. “Behaviors” represent the internal causal processes and mechanisms that enable the components’ functions. Hmelo-Silver and Pfeffer (2004) used this framework to examine students’ and experts’ representations of an aquatic system in terms of the parts, or the structural elements of the system, the elements’ behaviors or mechanisms, and the functional aspects of the system. They found that in contrast to the experts, students focused on the structures, providing little functional or mechanistic descriptions.

Mechanistic Reasoning. Mechanistic Reasoning (Bechtel & Abrahamsen, 2005; Glennan, 2002; Machamer, Darden, & Carver, 2000; Russ, 2006; and Russ et al., 2008) is an analytical lens that can be used to categorize a learner’s non-teleological, causal explanations of phenomena they have experienced. Mechanistic explanations focus on the processes that underlie cause–effect relationships and thereby take into account how the activities of the constituent components affect one another (Machamer, Darden, & Carver, 2000; Hammer et al., 2008; Russ et al., 2008; Bolger, Kobiela, Weinberg & Lehrer, 2012). Development of mechanistic reasoning is central to the development of scientific expertise (Hammer, Elby, Scherr, & Reddish. 2005; Metz, 2004; Russ et al., 2008; Nersessian, 2008) as well as integral in the development of expert-like thinking regarding the behavior of complex systems (Jacobson & Wilensky, 2006). As Machamer, Darden, & Carver, (2000) pointed out, scholars have argued for the importance of mechanisms in biology (Kauffman; 1971; Brandon, 1985; Crick 1988; Bechtel & Richardson, 1993; Burian, 1996). Typically, in biology, the notion of a "mechanism" has been analyzed in terms of the decomposition of "systems" into their "parts" and "interactions" (Wimsatt 1976; Bechtel and Richardson 1993).

Russ et al (2008) proposed a framework for analyzing mechanistic reasoning in student talk around a specific scientific phenomenon. This framework is comprised of seven categories of explanations. They are: describing the target phenomenon (Category 1); identifying enabling conditions of the environment that allow the phenomenon to take place (Category 2); identifying entities, activities and properties of the system (Categories 3, 4 and 5); identifying spatial and/or structural organization of entities (Category 6); and chaining (Category 7). During *chaining,* learners use knowledge about the causal structure of the phenomena to make claims about what must have happened previously to bring about the current state of things (*backward chaining*), or what will happen next given that certain entities or activities are present now (*forward chaining*). Russ et al. (2008) explained that the development of mechanistic reasoning can be understood as the process of development of learners' *sense of mechanism* (diSessa, 1993), i.e., the process by which learners try to identify salient events that constitute a phenomena, and which events follow which others. In this paper, we adopt the framework of mechanistic reasoning proposed by Russ et al. (2008) to trace the progressive deepening of students' explanations regarding the behaviors of agents and interactions between agents in the relevant ecosystem. A summary of Russ' coding scheme can be found in Table 1. We explain the reasons for adopting this *specific* framework in the sub-section below.

Table 1: *Summary of Mechanistic Reasoning Framework (Russ et al, 2008) and its application to student talk and written work in our study.*

Level	Category	Description of Category	Interpretation of mechanistic explanations in our study
1	**Describing the Target Phenomena (DTP)**	The learner states or demonstrates the particular phenomenon or result they are trying to explain	"I was trying to get energy by drinking nectar" "I needed to drink nectar from the flowers to stay alive"
2	**Identifying Setup Conditions (SC)**	Learner identifies particular enabling conditions of the environment that allow the mechanism to run.	"I started with 15 units of energy" "There were flowers around the room" "I have a proboscis"
3	**Identifying Entities (IE)**	Learner identifies objects that play a role in and affect the outcome of the phenomenon	"I had a [certain type, either short or long] proboscis" "I went to that flower" (flower location) "I didn't have much energy left"
4	**Identifying Activities (IA)**	Learner articulates actions and interactions that occur among entities; learner recognizes that actions of some entities cause changes in the surrounding entities	"I used my proboscis to drink from the flower" "I took 8 steps to get to that flower"
5	**Identifying Properties of Entities (IPE)**	Learner articulates general properties of entities that are necessary for this particular mechanism to run	"I had a long proboscis so I could drink from every flower" "I had a short proboscis so I could only drink from small flowers"
6	**Identifying Organization of Entities (IOE)**	Learner attends to how the entities are spatially organized, where they are located, and how they are structured.	"I went to a close flower that I could drink from to not lose much energy" "I had a long proboscis so I could drink from anything so I chose to drink from close flowers"
7	**Chaining (C)**	Learner uses knowledge about the causal structure of the phenomena to make claims about what must have happened previously to bring about the current state of things (backward) or what will happen next given that certain entities or activities are present now (forward).	"I set the flowers to clump so the butterflies don't have to fly as far" **Researcher:** "If you change the color of the butterflies to red, what will happen?" **Student:** "They (butterflies) will die"

*Affordances of Russ et al. (2008)'s Framework for our Study***.** We find Russ et al.'s (2008) framework of mechanistic reasoning to be well aligned with the paradigm of agent-based based computation. Pedagogically, the affordances of using mechanistic reasoning as a framework for knowledge analysis in our study are twofold. First, it can be used for characterizing the growth (i.e., the process of progressive refinement) of learners' reasoning about at least some key aspects of complex systems. Second it allows us to identify which kinds of mechanisms might be challenging for students. Both of these points are pedagogically important because it is challenging for novices to identify functional and behavioral explanations of complex systems (Hmelo-Silver et al., 2007; Hmelo-Silver & Pfeffer, 2004), and further, Hmelo-Silver & Eberbach (2014) have argued that in ecosystems, "the interrelationships between structural and behavioral/functional levels represent mechanistic explanations of ecological phenomena" (p. 411).

Along the first dimension, we posit that Russ' (2008) framework provides us with a more nuanced lens to identify elements of these interrelationships, and the process of development of these interrelationships. Note that Russ' framework explicitly requires the learner to focus on the agents or entities within the system (Category 3), the properties, actions and interactions between those agents and their corresponding events (Categories 4, 5 and 6), and the relationships between events in the system (Category 7). Consider, for example, the population dependence of two species in a predator-prey ecosystem. The pattern of dependence of the predator population with the prey population can be understood as a result of the combinations of simple, agent-level behaviors such as eating and movement (among others). These individual-level rules of interaction, the resultant "events", and relationships between events correspond to Russ' categories of mechanistic reasoning. Studies show that understanding individual-level rules and interactions involve understanding properties and actions of agents, as well as interactions between agents which, in turn, become meaningful events in the course of interpretive actions of the learner (Dickes & Sengupta, 2013). Therefore, this framework may allow us to specifically identify *how* students may identify and explain system level outcomes (e.g., population survival or demise) in terms of agent-level behaviors and interactions (e.g., camouflage and foraging).

Along the second point, identifying the components of mechanistic reasoning pertaining to different levels of the phenomenon will enable us to provide more specific guidance as to which *kinds of mechanisms* are challenging for students. For example, it may help us identify how students are struggling to appropriately identify the order of events and relationships between events that become evident as the simulation unfolds over time. This is particularly evident in our study in the context of students learning to identify the role of camouflage as an effective mechanism for survival. Although they do not specifically use the terminology of Russ' (2008) mechanistic explanations, Wilkerson-Jerde & Wilensky (2014) have similarly showed that attending to agent-behaviors and agent-level events (Levy & Wilensky, 2008), as well as identifying patterns of change in events and accumulations of these events, can provide novices (high school students) a deep understanding of system-level behaviors as they interact with multi-agent simulations of population dynamics.

Beyond the Single Mechanism: Reasoning about Loose Horizontal Coupling & Functional Equivalence.

Our description of mechanistic explanations thus far leaves out the important case of loosely coupled events and interactions (Simon, 1977) that we described earlier. Reasoning about loose coupling and functional equivalence requires the consideration of not only mechanisms within a single event, but also the consideration of alternate mechanisms or other events through which the same result can be obtained. That is, functional equivalence in our butterfly foraging system can be construed in terms of different classes of agents or sub-systems (i.e., long- and short-proboscis butterflies) each having the same effect (i.e., loss of nectar) on another class of agents or subsystem (i.e., short flower). Another example of functional equivalence can be understood in terms of the equivalence of both short and long flowers as possible food sources for the long proboscis butterfly. Thus, both long and short flowers are functionally equivalent to the long-proboscised butterflies, whereas the short proboscised butterflies can only drink nectar from short-stemmed flowers. This results in a distinct biological advantage for the long-proboscised butterflies. In this context, reasoning about functional equivalence involves reasoning about the variation of the structure, behavior and function of the relevant attributes among the agents, and their implications for the system-level outcomes such as population dynamics.

The Learning Environment

Our learning activities are divided into three phases, and are bookended by a pre- and post-assessment described below. For reference, Table 3 offers a summary of the sequence of activities as well as the relevant learning goals.

Phase 1: Embodied Modeling Activity

During this phase, students participated in an embodied modeling activity in which they acted as butterflies foraging for nectar. A photo of the classroom setup during this activity, and students conducting this activity is shown in Figure 1. Each student was given either a tall or short plastic straw that represented a butterfly's proboscis. Artificial flowers of two different stem lengths, short and tall, were placed in plastic jars throughout the classroom. Small glass beads represented energy in a discrete, tangible form. At the beginning of this activity, beads were placed in plastic cups to represent flower energy and each student was given a carrying pouch with fifteen beads to represent their starting energy. Students either deposited or collected beads to or from the plastic cups (i.e., flowers), depending on their foraging actions. Students gained 5 units of energy if they were able to drink nectar (i.e. their proboscis was able to ‘reach’ the nectar in the flowers) and lost one unit of energy for every step they had to take during foraging. At the end of every forage—once students reached the intended flower and collected beads from it—the student would calculate the *net* energy lost or gained on a printed data sheet by first subtracting from their total energy the energy it took to reach the intended flower and then adding back the energy they gained from drinking from it. Students’ actions during this activity represented “agent-level” rules of the ABMs that were later introduced in Phase 3, described below.

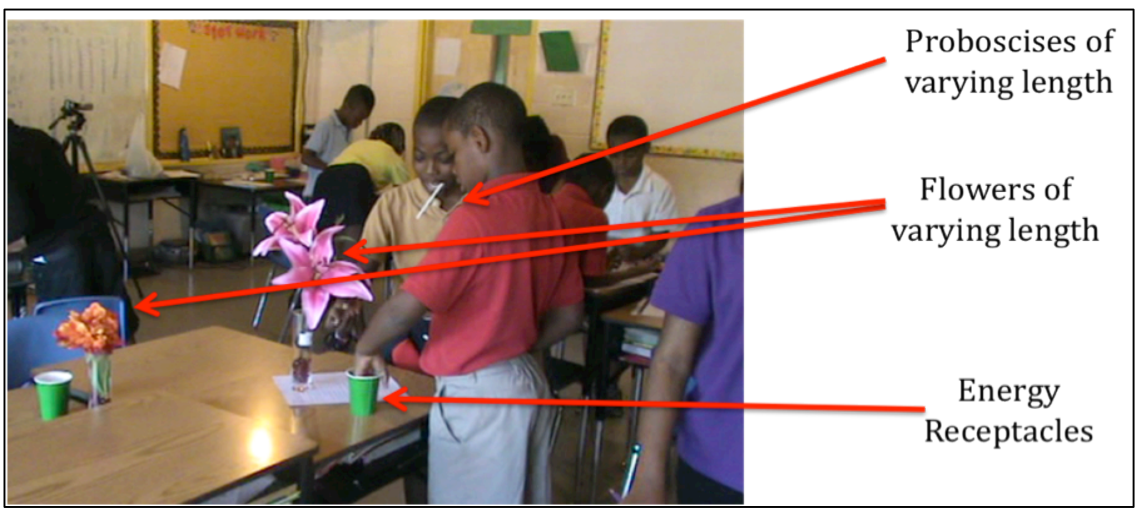


Figure 1: Students participating in Phase I’s Embodied Modeling Activity

Previous research by Danish and colleagues highlight the importance of constraining the interactions between students and the simulated physical environment during the embodied modeling activity. They found that this is essential in order to enable students to maintain focus on *key* aspects of the target phenomenon, relevant to the learning objectives. In Danish's work, the focal learning goals were to understand how bees communicate with each other, and how their communication affects emergent behaviors of hives around nectar. In an early version of this work, the goal for students acting as "bees" during the embodied modeling activity was to collect as much nectar as possible (Peppler et al., 2010), however, the authors found that the students cheated to win by collecting nectar as quickly as possible, instead of spending adequate time on communicating with other bees. As Danish (2014) pointed out, remedying this situation involved redesigning the activity by changing the rules of the game, so that students worked in pairs to hide nectar and then create a dance to communicate the nectar location to their peers, who would then search for the indicated location. That is, the activity more explicitly highlighted communicating with bees—the target learning goal—as an agent (or student)-level action that was *necessary* in order to complete the game.

Based on this body of work, we designed actions, performed by students-as-agents, interwoven with reflection that supported the intended learning goals of familiarizing students with the various "agent-level" elements of the ecosystem such as flower location, depth of the nectar sacs, and proboscis length (Table 2). Some of the rules and actions (e.g., losing energy due to travel, and gaining energy due to food intake) were designed to leverage students' intuitive understanding of the relationship of energy and physical activity. Reflection was woven into the activity through the form of energy data sheets, wherein students actively recorded their steps taken (energy lost) and their nectar intake (energy gained) to maintain a running total of their change in energy over time.

Table 2: Agent-level Rules and Variables Introduced to Students in the Embodied Modeling Activities

Variable	Rules of the System
Energy	1. Each step toward a flower costs one unit of energy 2. Each flower gives 5 units of energy
Flower Location	1. Far away flowers cost more energy 2. Close flowers cost less energy
Flower Length	1. Nectar in tall flowers is difficult to collect 2. Nectar in short flowers is easy to collect
Proboscis Length	1. Short Proboscises can only drink from short flowers 2. Long proboscises can drink from any flower

It is important to note that during the embodied modeling activity, the need to move from embodied actions to recording vital information on a data sheet did not disrupt students' immersive experience as agents within the system. Students recorded the relevant data as it happened, and it did not require students to give up their butterfly persona, although some students found jotting and foraging at the same time initially disruptive. The teacher and the researchers strongly believed that interweaving action and reflection was central to achieving the learning objectives of thinking like an agent. Therefore, to help students who expressed such difficulties, they explained that if the student did not know how much energy she or he had at the beginning of each forage, she or he would not be able to make decisions about foraging appropriately, and this could in turn result in death. This clarification helped frame the jottings of their energy gains and losses as an integral part of the embodied modeling experience. Students completed two separate iterations of this activity. After each iteration students created bar graphs of their energy, described in the next section. In the second iteration, each student was provided with a straw of a different length (compared to the first day), and was asked to begin their forage from a different starting position. By changing their initial starting location, our goal was to make sure that students did not repeat their actions from the previous iteration. This, in turn, created opportunities for the students to reflect on the differences between the two iterations (see the section on Teacher-Researcher Partnerships), and necessitated a deeper engagement with the simulated physical environment.

Phase 2: Generating Foraging Maps and Bar Graphs of Energy During Foraging

After completing each iteration of the embodied modeling activity, students used their *total energy* data recorded on the energy data sheets to create representations of how their energy changed over time (Figure 2a). Students were also provided with maps of the classroom with the locations of the flowers identified, which they annotated by marking their foraging paths (Figure 2b). The researchers asked students to "show us how your energy changed over time using the materials we provided (construction paper and paper strips) and your own materials (markers, glue, stickers, etc)." The classroom teacher, dissatisfied with the ill-defined nature of the task, subsequently instructed students to 'make a bar graph' of their energy. This and other instructional constraints are discussed in more detail in the section on Researcher-Teacher Partnerships.

Despite these constraints, the bar graphs and the foraging maps gave students an opportunity to represent and explain change over time by accumulating discrete events and comparing those events to their embodied actions as agents within the system. The rules for calculating energy were designed to prompt students to think about what each embodied action represented in terms of energy gains and losses, whereas the overall graph made explicit the change in energy as a result of accumulation (over time) of many steps, grouped in the form of forages. Reasoning about such patterns of accumulation is an important aspect of learning about aggregate-level outcomes using ABMs (Wilkerson-Jerde & Wilensky, 2014).

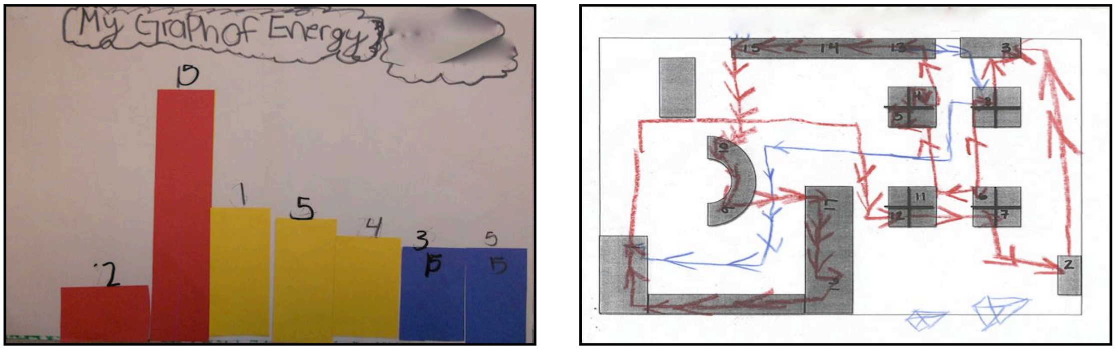


Figure 2: Sample student generated energy graphs (left – Fig. 2a) and foraging maps (Right – Fig. 2b).

Phase 3: Inquiry with ABMs

During this phase, students interacted with two agent-based computational models designed in the NetLogo modeling platform (Wilensky, 1999). The computational rules obeyed by the agents in the ABMs were identical to the rules students followed as butterfly-agents during the embodied modeling activity. Overall, our goal was to encourage students to use their prior experiences with the embodied modeling and graphing activities to scaffold their inquiry with the ABMs. The first ABM students encountered, Model 1, simulated predator-prey dynamics in an ecosystem of flowers, butterflies and birds (Figure 3a). Three of the variables (proboscis length, flower length and flower location) were familiar to the students based on their embodied modeling and graphing experiences. In addition, this model also introduced the following variables: 1) color of flower, 2) color of butterfly and 3) predation. The goal of this activity was to identify the model parameters that resulted in a thriving butterfly population (see section on Researcher-Teacher Partnerships). Successfully creating conditions in the simulation for a thriving butterfly population required an understanding of the role of camouflage on butterfly survival, i.e. butterflies whose colors were closer to that of the flowers had greater chances of survival. The introduction of new variables into the system was designed to progressively increase the number of variables, as well as introduce aggregate-level behaviors that would be difficult to recreate in the embodied modeling activity, such as population change over time. In this case, the added complexity of Model 1 enabled students to investigate a new phenomenon (predation) and test ‘what-if’ scenarios based on both previously encountered outcomes and as yet unimagined outcomes.

In addition to all the components of Model 1, Model 2 (Figure 3b) provided students with an option to highlight the behavior of a single agent (butterfly) during the simulation. When students highlighted the behavior of a single butterfly in the model, they were also able to see a line graph of that butterfly’s energy vs. time (see Figure 3b). This graph was similar to students’ paper-based bar graphs of energy generated during Phase II. In order to understand the graph in the simulation, students had to consider what was happening to the energy of the butterfly during every “step” of the computational model. The pedagogical objective here was to support students in bootstrapping their bar graphs of energy in order to meaningfully interpret a continuous line graph representing change over time.

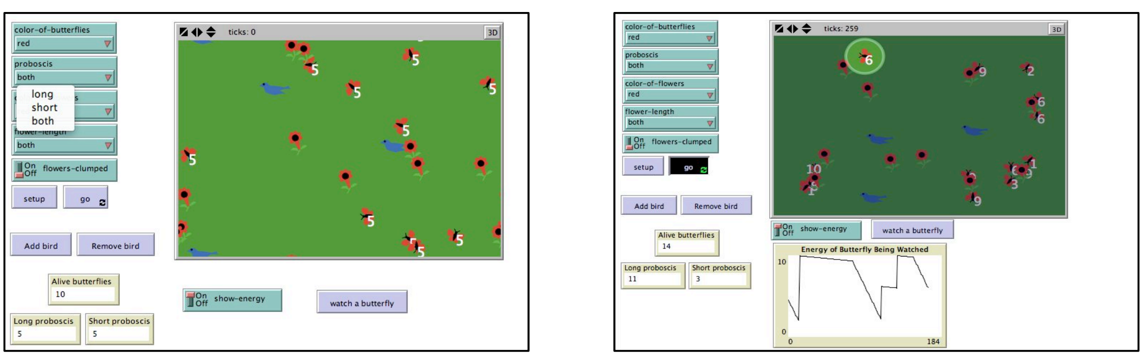


Figure 3: Screenshots of the Predator ABM (left) and Watched Energy ABM (right)

Pre- and Post-Assessments

During both the pre- and post assessments, students were provided large sheets of construction paper and images of various plants and animals. The goal for students was to design an ecosystem of animals and plants, and to represent the producer-consumer relationships between the various components of the ecosystem. Student ecosystems included both biological agents such as animals and plants, and environmental agents such as the sun and bodies of water. Students were allowed to draw their own agents if they preferred. During both the pre- and post-assessment, students were verbally prompted by the classroom teacher to think about what each animal in their system would need to survive and to think about how energy might flow through their system (e.g. from producer to consumer). Students depicted the flow of energy as links, either an arrow or a line, between agents in the system. Examples of student pre- and post-assessments can be found in Appendix B and C.

Researcher-Teacher Partnership & Instructional Moves

A central goal of our research is to develop activities that can be easily integrated with existing elementary science curricula. Such integration involves not only aligning our curricula with state and local 3rd grade ecology standards and but also working with the school principle and classroom teacher to make sure that institutional constraints are met. The specific state performance indicators, (SPIs) from the local standards documents, that guided our activity design are listed below.

1. Select an investigation that could be used to answer a specific question.
2. Investigate an organism's characteristics and evaluate how these features enable it to survive in a particular environment.
3. Investigate populations of different organisms and classify them as thriving, threatened, endangered, or extinct.
4. Identify the basic needs of plants and animals.
5. Recognize that animals obtain their food by eating plants and other animals.

Classroom instruction was shared between the first author and the classroom teacher, however, the classroom teacher was instrumental in guiding and constraining learning objectives in different activities, based on her interpretation of what was most "needed" (to quote her) to both work within the stated SPIs, as well as to maintain continuity with her previous instructional foci. While in some cases this meant diverging from activities initially planned by the researchers, such divergences are essential for maintaining teacher agency within the researcher-teacher relationship. In our discussion of instructional moves that follows, we explain how the teacher's perspective shaped and constrained the learning goals and activities.

Instructional Moves in Phases 1, 2 & 3

During Phases 1 and 2, the teacher and the researcher circulated throughout the classroom and assisted students with any questions they had. The classroom teacher was concerned with students "getting the math correct" during the activity while the primary researcher was more open to students remaining fully immersed in the activity. As we mentioned earlier, the inclusion of the energy data sheets did not remove students from the space of activity and we have no evidence that indicates that the classroom teacher stopping to assist students with mathematical errors disrupted their flow of activity and ideas. After completing the embodied modeling tasks, students were instructed to create representations that depicted how their energy changed over time, later constrained by the teacher to include only canonical bar graphs. The teacher explained to the researchers that her decision to limit the forms of representation to bar graphs was based on the existing third grade math and science standards. Bar graphs are ubiquitous features of high stakes assessments across the elementary grades in this state, and the classroom teacher also saw this activity as an opportunity to reinforce earlier instruction about bar graphs. This decision greatly limited opportunities for students to create and refine their own "invented" representations. Thus, although students' representations and explanations became progressively more complex, the prescribed nature of the mathematical representations limited the potential for what Enyedy (2005) termed *progressive symbolization* of their own, invented representations.

During both iterations, the first author conducted informal interviews with students, asking them to explain how their graphs represented their data and how their actions during each forage were represented in each bar of their bar graphs. Students were also asked to explain how their graphs corresponded to their foraging maps. Following the second iteration, students were also asked how their graphs and foraging paths had changed across both iterations and to provide written explanations about which path was more advantageous and how they, as butterflies, made decisions to increase their survival. These independent work sessions were followed by a class discussion where students shared their collective foraging ideas, talked about why they made certain decisions, and thought about how those decisions were influenced by their proboscis lengths, depth of nectar sacs and the location of flowers. During this discussion, students also spoke about which iteration was better and why.

During Phase 3, the teacher and the researcher circulated throughout the classroom and had conversations with pairs of students working next to each other. The goal of these conversations was to ask students to think aloud, both with their neighbor as well as the teacher, about their reasoning pertaining to the questions on the worksheet. During their investigations with Model 1, the classroom teacher also circulated among students, asking students to explain aloud what effect each of the variables had in terms of outputs and to also explain what effect these outcomes had on the population in terms of survival. In their worksheets, students were asked to provide three types of explanations: a) Explanations of the *population-level* outcomes of their simulations when they selected different parameters (e.g., why all the butterflies died or some survived when certain conditions were selected in the model); b) Explanations of the energy graph (in Model 2) in terms of relevant behaviors of the *agents* in the simulation; and c) Explanations of how similar or different the simulated graph was compared to their bar graphs during Phase 2. The prompts for explanation types A and B were specifically designed to engage students in multi-level reasoning (i.e., explaining aggregate level outcomes in terms of agent-level interactions). The prompt for explanation type C was designed to bootstrap students embodied (i.e., agent-level) experiences during Phases 1 and 2.

In addition, our field notes also indicate that during their interactions with Model 2, all the students spent a considerable amount of time observing the behavior of several different butterflies using the "Watch a butterfly" tool. Most students began using this tool on their own, but a few students were prompted by the teacher to begin using it. During their visits to the students, the researcher and the teacher asked the students to explain the graph of one butterfly, as well as to explain why different butterflies, when observed individually, showed different graphs.

In Phase 3, the teacher also guided the activity in a manner that led students to focus on population growth and survival as the key system-level outcome. This instructional move was based on her interpretation of the state standards for ecology. This in turn resulted in her scaffolding the students to reason about the mechanisms behind thriving and/or dying populations. The emphasis on population survival came as a tradeoff with creating opportunities for comparisons across subsystems, such as competition and loosely coupled events. The latter could have brought to the foreground the role of chance, variation and the distribution of agents and resources within the system, and in subsequent studies we have addressed this issue by developing agent-based modeling activities that integrate with 3rd grade science curriculum over extended periods of time (Dickes et al., 2015).

Research Questions

In this study, we investigated the following research questions:

1. How do students develop mechanistic explanations of ecological phenomena as they progress through Phases I and II?
2. How do students' embodied modeling and representational experiences in Phases I and II shape their interactions with the ABMs in Phase III?
3. What are the forms of reasoning about the following elements of complexity?
    - *3.1* Loose coupling
    - *3.2* Population growth and survival
    - *3.3* Inter-agent and agent-environment relationships in food chains

Method

Participants and Setting

The study was conducted in a 3rd grade classroom consisting of 17 students in a 98% African-American public charter school located in a large metropolitan school district in the southeastern United States. All students in the class were eligible for the federal free lunch program. All students performed below grade-level in at least one area on standardized assessments earlier in their 3rd grade academic year.

The study took place during the regular science class period. Data was collected over a period of two weeks and included 7 days of activity. The duration of activity was about one hour and 15 minutes for each of the seven days of the study. Two students were absent for more than half of the duration of the study, resulting in an effective sample size of 15 students. Our lessons did not replace the students' normal science class; rather, they were designed to complement third grade ecology curriculum. A summary of the instructional schedule, the time spent on each activity and the relevant learning goals is provided in the Appendix A.

Data and Analysis

Forms of Data. The data for the selected case and classroom level analysis comes from informal interviews with the participants, video recordings of class activities and discussion, student artifacts (e.g. student energy graphs, maps and worksheets and pre- and post-assessments) and field notes. Informal interviews were conducted by the first author during opportune moments while the students were engaged in the modeling and representational activities. A second researcher who was also present every day in the classroom videotaped these interviews.

In some cases, the interviews ensued when the student called upon the teacher or researcher in order to help him or her with a difficulty. In other cases, interviews were conducted in order to ask students to explain their thinking and reasoning about the different inscriptions used in the study. Throughout the study, during every class, we first interviewed four focal students – two of them high performing, and two of them low performing in their regular science class, as identified by the teacher. After we interviewed these students, time permitting, we then interviewed typically 3 to 6 additional students.

Case Study Approach. We present the analysis in the form of explanatory case studies (Yin, 1994). As Yin pointed out, *explanatory* case studies are well suited as a methodology to answer *how* and *why* questions. One of our goals is to illustrate the *process* through which students developed mechanistic explanations; in other words, *how* their embodied and representational experiences shaped their interactions with the ABMs and *why* certain forms of reasoning about complexity are more difficult than others. Following previous qualitative studies in science education focused on identifying the processes of students' conceptual development during a curricular sequence (Petri & Niedderer, 1998; Taber, 2008; Dickes & Sengupta, 2013), our selection and analysis of cases were guided by the following two criteria: representativeness and typicality.

Representativeness implies that the selected cases should aptly represent key aspects of learning experienced by the students. These key aspects or themes, in turn, are defined based on the research questions. In our study, the criterion of representativeness implies that the selected case should highlight the relevant representational practices and conceptual understandings as defined by our research questions. The representative themes pertaining to each research question are presented later in this section. In order to identify representative cases, we looked at several factors along different dimensions: *logistical* (Was the student present during every class?), *technical* (Does the video and audio quality of the particular interview or excerpt of class discussion recorded in the video allow for a transcription of all of the salient elements of the conversation?), and *theoretical salience* (Upon analysis, how clearly can the recorded conversation highlight the relevant theme?). *Typicality* implies that the selected case(s) should potentially offer insights that are likely to have wider relevance for the remainder of the participants in the study. In other words, the cases selected should represent aspects of the process of learning experienced by a majority of the student population. Pertaining to each theme, we present the individual cases, as well as a classroom-level analysis of relevant responses of all the students. The classroom-level analysis provides evidence about the typicality of the responses evident in the respective individual cases.

In order to identify representative themes, we used the check coding method (Huberman & Miles, 1994). In this method, two or more researchers independently code data and then clarify their differences until consensus is reached. This work was conducted in three phases. A first pass at data analysis was conducted jointly with the first and second authors. We each watched the videos of our assigned students and noted segments that seemingly related to explanations of conceptual thinking. We recorded our initial observations and discussed these as a group with other members of our research lab. These observations were mainly descriptive in nature, and corresponded to what Huberman and Miles (1994) term *descriptive codes*. We identified sets of distinct descriptive and pattern codes pertaining to each research question. For example, for RQ1, the descriptive code for Dontavia's explanation in Excerpt E1 was "Dontavia explains her actions and behaviors as a butterfly". After this initial pass, transcriptions of all the interviews, class discussions, and written responses were generated for all the cases selected for analysis. We then began open coding (Corbin & Strauss, 1990). During this phase of analysis, we carefully re-watched the interview videos and re-read the transcripts, with the goal of generating analytic codes that Huberman and Miles (1994) term *pattern codes*. A pattern code is inferential, a sort of meta-code, that pulls together the data labeled by descriptive codes into smaller and more meaningful units—this implies that for each descriptive code, there were multiple pattern codes. For example, for RQ1, the pattern codes for Dontavia's explanation in Excerpt E1 involved identifying the specific elements of mechanistic reasoning, based on Russ et al.'s (2008) coding scheme (e.g., *setup conditions, entities,* etc.) as evident in her statement.

We follow the trajectory of a student – Dontavia (pseudonym) – throughout the course of the learning activities, and use her case to present our analysis of RQ1 and RQ2. We chose Dontavia based on the criteria of representativeness and typicality. The criterion of representativeness was applied as follows: Dontavia was present every day during the study, and was interviewed regularly during the study. Furthermore, in comparison to other interviewees, Dontavia's responses were more detailed during every interview. All of her video recordings consistently met our logistical, technical and theoretical salience criteria for representativeness. The criterion of typicality was applied as follows: Dontavia's trajectory represents the typical learning experiences of the majority of the students, as evident in our comparisons of Dontavia's work with the interviews and written responses of other students, which we also present in our analysis.

Analysis for RQ1. Our first research question analyzes the development of mechanistic explanations across Phases I and II of the learning sequence. In order to answer RQ1, we analyzed informal student interviews and students' material artifacts generated during Phases 1 and 2 using the coding scheme of mechanistic reasoning (Russ et al., 2008). The data for this analysis includes students' explanations of how their energy changed during each iteration of foraging, as represented in their energy data sheets, bar graphs of energy consumption during foraging, and their foraging maps. These explanations were both in the form of written responses (all students) and videotaped interviews conducted with a smaller number of students who were followed throughout the unit. After completing each iteration of the embodied modeling and graphing activity, students were asked to provide written explanations of how their foraging actions were represented in the graphs. Table 1 provides Russ et al.'s (2008) coding scheme as well as offers examples of how we interpreted the different levels of mechanistic explanations in student talk and written work.

Analysis for RQ2. Our second research question investigates how students' embodied modeling and representational experiences shaped their interactions with the ABMs. In order to answer RQ2, we analyzed student talk during the representational activities as well as during their interaction with the ABMs. Two themes emerged in our analysis: the use of egocentric explanations in which students projected themselves as the simulated agents, and the further sophistication of students' mechanistic reasoning in the form of *backward* or *forward chaining* (Level 7 mechanistic explanations). Backward or forward chaining was evident in the form of the order in which students explored the variable space in the simulations – i.e., they first manipulated known variables within the system to test 'what if' questions, and then explored variables and mechanisms that had not been previously encountered. The data for this analysis comes from students' written work and informal interview data, which made explicit the order in which they controlled the variables of the simulation in Model 1 in order to produce thriving populations of butterflies. We also conducted interviews with several students *during* their interactions with the computational models. In these interviews, we asked them to explain their strategies for keeping the butterflies alive in Model 1. Data from these interviews was compared with students' written explanations to determine the order of variable exploration in Model 1. During their explorations with Model 2, students were asked to explain, both on activity sheets and in informal interviews, the similarities and differences between the simulated graphs of Model 2 and their graphs of energy change they constructed during the embodied foraging activity. Students' egocentric explanations were also evident during these interviews, as they projected themselves to be the butterflies on screen they were attending to.

Analysis for RQ3. Our final research question explores the forms of reasoning about some of the key characteristics of complexity in the system, specifically in the form of loosely coupled events, survival and population growth, and interdependence. In order to analyze student understandings of loosely coupled events within the system, we coded student talk during both individual interviews and classroom discussions for the presence of statements indicating functional equivalence and ecological advantage (e.g. "I could drink from any flower because I had a long proboscis."). In order to analyze student thinking around population survival, student interview data was coded for explanations of the role of camouflage in population change. Specifically, we coded for explanations that were able to identify that camouflage helped population increase. Finally, to analyze student thinking about interdependence, we coded students' pre- and post-assessments in terms of the number of *links* between agents that represented the direction of energy flow (i.e., producer-consumer relationships). Students depicted links as either arrows or lines from one agent to another and were counted accordingly. Sample coded pre- and post-tests are provided in Figures 4a and 4b.

Reliability. The codes for qualitative analysis of the interview data as well as the written pre- and post-test data were first developed through mutual discussion between the lead authors of this paper. The first and second authors jointly coded the data over a period of several months, and through regular discussion grounded in the theoretical framework discussed previously, came to agreement about development and application of the codes. Once agreement between the two authors was reached, this analysis was shared with the remaining authors, and further revised based on their feedback. The resultant analysis reflected agreement among all the authors. After this, the data presented in this paper was blind-coded by a researcher, Paula, who was unaffiliated with the study. Paula is a researcher in social sciences with extensive background in qualitative analysis. Paula coded all of the data presented in this paper, and agreed with the authors' codes an overall 86.67% of the time, resulting in a Cohen's Kappa of 0.83.

Findings

RQ 1: Development of Mechanistic Explanations Across Phases I and II

The first portion of our analysis explores the development of mechanistic explanations during Phase I and II of the learning sequence. We first trace the development of mechanistic explanations in one representative student, Dontavia, and follow with descriptions of classroom level findings.

Dontavia's Case. In Excerpt 1 below, Dontavia explains the actions she took that resulted in her foraging map as well as her energy graph. She identifies distance travelled – i.e., the number of steps needed to reach a flower - as the most important mechanism of energy loss. Dontavia's selection of distance travelled as the most important mechanism for survival is reflected in her foraging map which reveals that she visited only *proximal* flowers. As evident in lines 2 – 7 (Excerpt 1), she travelled to the closest flowers, regardless of whether her proboscis allowed her to drink from that flower. At first, Dontavia's energy level increased with each forage for the first few acts of foraging; however, her energy levels took a noticeable dive on her last two forages because she chose to travel to the most proximal flower without considering whether or not she could actually drink from it. As a result, her movement cost her a considerable amount of energy that she could not recuperate since she couldn't drink from the flower. This is evident in her energy data and spatial map of foraging, shown in Figures 7 and 8 respectively. In visiting flowers next to one another, Dontavia's overall goal was to travel to all the flowers in the room, as evident in Line 9, Excerpt 1.

Excerpt 1:

1	Researcher	So, talk me through what you did here [on your energy data sheet].
2	Dontavia	I started with my flower that was at my desk first and then I took
3		five steps and went over
4		to that [closest] flower and got energy and then I took 4 steps and I

5		went over to that
6		[closest] flower and then I took another five steps and went over to
7		that flower there.
8	Researcher	Why did you choose these flowers?
9	Dontavia	I was just trying to get to all of the flowers.

We will now examine the categories of mechanistic reasoning evident in her work. Table 1 shows descriptions of these categories. From her responses, it is clear that Dontavia was aware of her goal, which was to forage in order to gain energy by drinking from flowers (line 9, Excerpt 1). This indicates that she was aware of the *target phenomenon* (Category 1) that she was modeling. Her interview responses, and her energy data sheet also implied that she was aware of an important *setup condition* (Category 2): i.e., having a proboscis allowed her to drink from flowers. Dontavia was also able to explain that both flower location, i.e., *entities* (Category 3) and the number of steps she took to get to that flower, i.e., *activities* (Category 4), affected her energy. However, she did not consider her proboscis length as a factor in deciding which flowers to visit; instead, her foraging decision was guided by the goal of "just trying to get to all of the flowers", indicating that the goal of the activity was the most important factor in her decisions about where and how to forage. A summary of Dontavia's mechanistic explanations during Iteration 1 is provided in Figure 4.

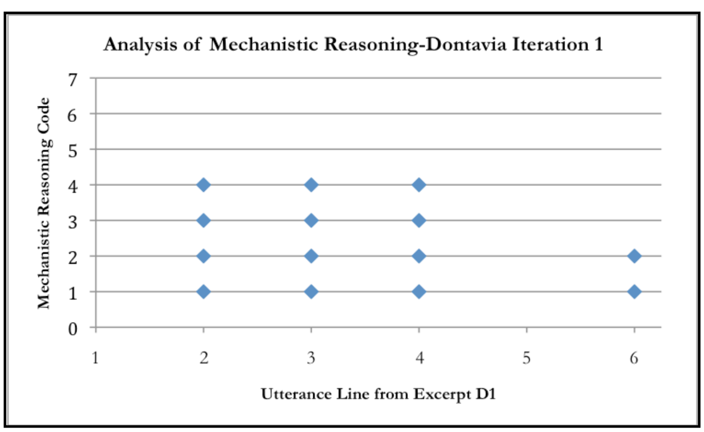


Figure 4: Analysis of Dontavia's Mechanistic Reasoning during Iteration 1

Dontavia's second attempt at foraging for nectar was quite different than her first. She began to make foraging decisions based on her proboscis length – i.e., by reasoning about factors beyond the stated activity goal. As shown in Excerpt 2, she commented that after an unsuccessful foraging attempt, she went to a flower that "I *could* [emphasis added] drink from" and her energy went up. The use of the word "could" implies that she had now considered the match between her proboscis length and the flower's nectar depth as a factor in deciding which flowers to visit. Dontavia explained to a researcher that after a forage, she would look around her current location and try to visually observe the stem lengths of the visible flowers, before making a decision about which flower to visit next. Her statement that her energy kept getting "bigger and bigger" demonstrates that she continued to go to flowers that she was able to drink from, which is also evidenced in line 6 of the transcript. Furthermore, she also visited flowers that were proximal to one another, i.e., within a few steps of each other. She commented in line 8 that butterflies can waste their energy by walking needlessly, even if the flower is a flower the butterfly can drink from. When the researcher restated Dontavia's explanation of 'wasting energy through walking' (line 11), Dontavia interrupted the researcher (line 14) and pointed out that her strategy during the activity was to visit flowers that were only "one step apart". This indicates that not only did she choose flowers that matched her proboscis length, but that the flowers were close by, in order to avoid wasting energy. This is also evidenced in the form of her foraging path (Figure 7, right), which shows that Dontavia primarily chose to travel to flowers that were both close to her present location and commensurate with her proboscis length.

Figure 6: Dontavia's 1st (left) and 2nd (Right) Iteration Energy Graphs

Compared to Iteration 1, her average energy was higher and more stable, as evidenced in her energy graph (Figure 6). It is worth noting that while her energy still took a plunge during her final forage because she chose to travel to a far away flower; she only made this decision after she had accumulated enough energy to offset the energy loss, and therefore did not "die" due to complete energy loss.

Excerpt 2:

Researcher	Talk to me about what you did.
Dontavia	At first I had 15, and then I lost some 'cause I couldn't drink from it and then I went to another flower and I could drink from it and it got bigger and bigger and bigger.
Researcher	You said you lost energy because you went to a flower you couldn't drink from. Why did your energy keep getting higher and higher after that?
Dontavia	I could drink from those flowers.
Researcher	Is it possible to still lose energy even if you can drink from a flower?
Dontavia	Yes, if you waste energy by walking.
Researcher	That's right, if you waste energy by walking a long ways you can still lose energy//
Dontavia	//No, these flowers that I went to were only one step apart.
Researcher	Oh, okay. So did you go to close flowers that you could drink from to save energy?
Dontavia	Yes

Dontavia's explanations during Iteration 2 demonstrate the inclusion of new categories of mechanistic reasoning. Similar to her responses in Iteration 1, she continued to demonstrate an understanding of the *entities* (Category 3), *activities* (Category 4) and *setup conditions* (Category 2) of the *target phenomena* (Category 1). However, in addition, she also explicitly identified *general properties of entities* (i.e., Category 5) that were necessary for this particular mechanism (i.e., a successful forage): proboscis length of butterflies and nectar depth of flowers. Her comment in line 10 indicates that her decisions regarding where to forage were based on a coordination of multiple factors: her current energy, flower size, proboscis length and the approximate distance to the next flower. This further demonstrates that compared to her explanations in Iteration 1, Dontavia had developed a more sophisticated understanding of the relationships between agents and structures in the ecosystem, what Russ et al (2008) term an "Identification of the Organization of Entities" (Category 6). A summary of Dontavia's mechanistic explanations during Iteration 2 is provided in Figure 5.

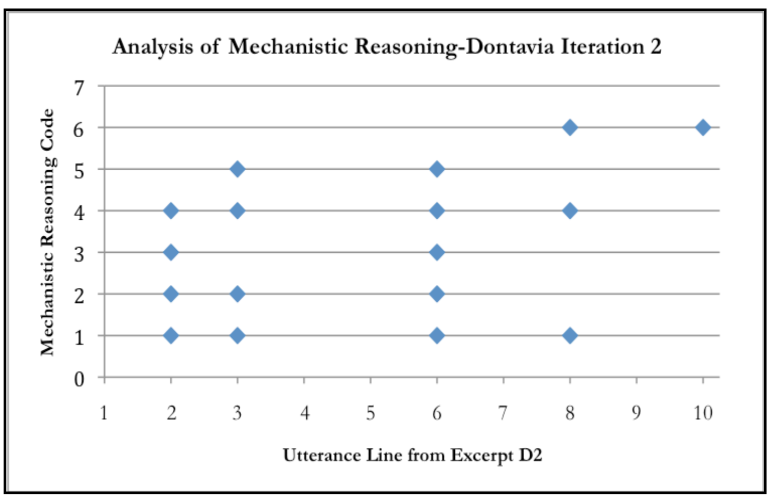


Figure 5: Analysis of Dontavia's Mechanistic Reasoning during Iteration 2

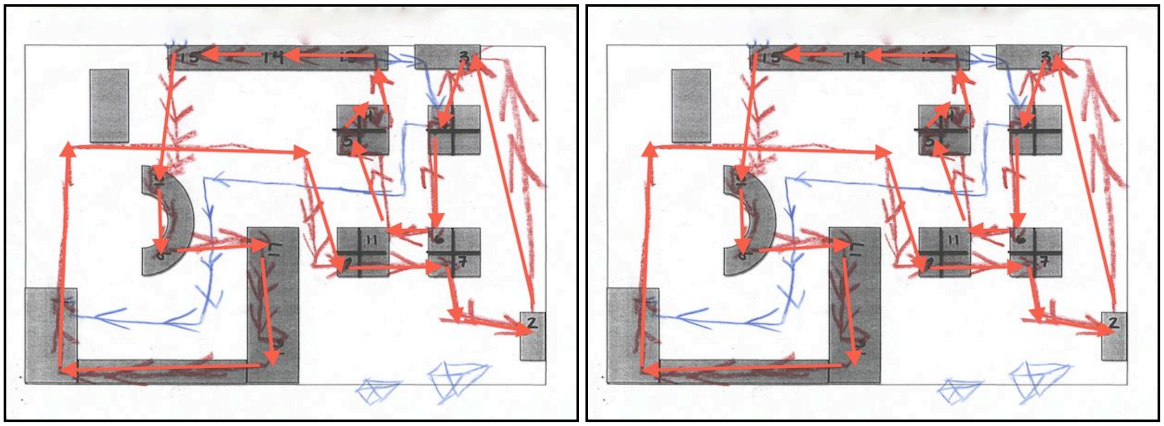


Figure 7: Dontavia's 1st Iteration (Left) and 2nd Iteration (Right) Foraging Maps

Classroom Level Findings. We found that similar to Dontavia, all students demonstrated an understanding of the target phenomena and setup conditions of the activity (Categories 1 and 2 in Table 1) during both the iterations. This is unsurprising given the rules of our embodied modeling activity. Before the first iteration of the modeling activity, students were given explicit instructions on how to successfully complete the activity, specifically how they lost energy (walking) and how they gained energy (drinking from flowers). By necessity, our instruction included both explicit and implicit information about the Target Phenomenon (foraging for nectar) and the Setup Conditions (proboscis length and flower length). As such, these lower level categories were 'givens' of the activity design. In contrast, the number of students who demonstrated Category 3 and 4 explanations in the first iteration was 93% and 73% respectively. This would suggest that these explanations were not 'givens' of the activity design but rather emerged through the course of engaging in the activity. The fact that Category 3 and 4 explanations had both increased to 100% of the class by Iteration 2 provides some support for this claim.

Table 3: Variables Identified by Students as Important to Survival and their relation to mechanistic explanations (N=15)

Category of Mechanistic Explanation	Iteration 1: Number of Students who demonstrated this level of thinking	Iteration 2: Number of Students who demonstrated this level of thinking
1 – TCP	15 (100%)	15 (100%)
2 – SC	15 (100%)	15 (100%)
3 – IE	14 (93.3%)	15 (100%)
4 – IA	11 (73%)	15 (100%)
5 – IPE	0 (0%)	9 (60%)
6 – IOE	0 (0%)	9 (60%)
7 – C	0 (0%)	0 (0%)

Also, similar to Dontavia's case, we found evidence of Category 5 (60%) and 6 (60%) explanations in students' interviews and written work in Iteration 2, but not in Iteration 1. Similarly, we found that Category 4 explanations were the highest level of explanations achieved by any student during Iteration 1. Category 5 and 6 explanations require a coordination of activities and entities within the system, relationships that were strengthened after second iteration of activity. Overall, these findings indicate the following: a) Students were productively using the knowledge they had gained during Iteration 1 in their work during Iteration 2; and b) they had begun to develop a more sophisticated understanding of the phenomenon by Iteration 2. A summary of the type and prevalence of specific mechanistic explanations across all students is provided in Table 3.

RQ 2: How students' embodied experiences in Phases 1 and 2 shaped their interactions with the ABMs in Phase 3.

Dontavia's Case. In Phase 3, during her interaction with Model 1, Dontavia was not initially successful in keeping the butterflies alive. The interview reported in Excerpt 3 took place after she had interacted with the model for nearly 20 minutes. During that time, she had designed and tested two different ‘what if’ experiments to keep the butterflies alive, varying the proboscis lengths of butterflies in each condition. During each of these conditions, the flowers were randomly arranged (spatially) within the NetLogo microworld.

Excerpt 3:

Researcher	How could you make the butterflies survive?
Dontavia	I’m going to change it to ‘both’ and clump the flowers. [Runs the simulation]
Researcher	Did the butterflies survive?
Dontavia	No.
Researcher	Why did the butterflies not survive? What else could you change?
Dontavia	[hesitates] Color of flowers? [pause] Oh, I know! [Changes flower color to red]
Researcher	Why did you choose red [flower color]?
Dontavia	Because they the same color.
Researcher	Why is that important?
Dontavia	[Pause] Because they camouflaged!

After these unsuccessful attempts, she told the researcher that in her model, she was going to “change the settings [in her simulation] to ‘both’” (line 2), in order to create a simulated ecosystem that includes butterflies with both long and short proboscises, as well as both tall and short flowers. This ensured availability of flowers to drink from for both types of butterflies. She also decided to “clump the flowers” (line 2) by altering the relevant setting in her simulation. Changing the flowers to ‘clumped’ reduces the energy cost for some butterflies in the system as it spatially arranges the flowers in close proximity to one another. In her written response, Dontavia explained that clumping the flowers would result in lower energy loss of butterflies due to foraging.

Next, Dontavia ran the simulation with these settings, and noticed that the butterflies eventually died (line 5). When the researcher asked her to think about what she could do to keep the butterflies alive, after a brief hesitation, she proposed that a new variable, flower color, may be a potential factor (line 8). In this line, her voice inflection indicated a brief moment of hesitation about flower color. After a quiet reflection of a few seconds, she realized the importance of butterfly color matching flower color, and changed the color of flowers (in the simulation) to red. The researcher asked her why she chose red, and Dontavia responded that the butterflies would be camouflaged because they were now the same color as the flowers (line 13). This excerpt demonstrates that Dontavia conducted her inquiry with the simulation in two phases. First, she productively used all the variables that she had learned to be important for a butterfly's survival based on her embodied modeling experiences: 1) Flower location, 2) Flower Type and 3) Proboscis Length. It was only after she found that these variables were unable to prevent butterfly extinction, that she considered and tested a new mechanism, camouflage, for its role in survival. This suggests that Dontavia used her previous experience of modeling foraging in an embodied manner in order to systematically explore the various familiar and unfamiliar factors and parameters in her simulation.

What does this mean in terms of mechanistic reasoning? Dontavia was using knowledge about the causal structure of the phenomena to make claims about what must have happened previously to bring about the current state of things and what might happen given the presence of new entities (butterfly and flower color). According to Russ et al (2008), this is evidence of *backward and forward chaining* (Category 7 in Table 1). For example, when her first attempt did not succeed, she again used her knowledge about the causal structure of the phenomena to make a new claim about would happen in the model given the new parameters she had set (line 6). Dontavia then proceeded to infer a causal mechanism to her explanation of the phenomenon (line 13): camouflage (or blending in) leads to survival. A summary of Dontavia's mechanistic reasoning during her interaction with the first model is provided in Figure 8.

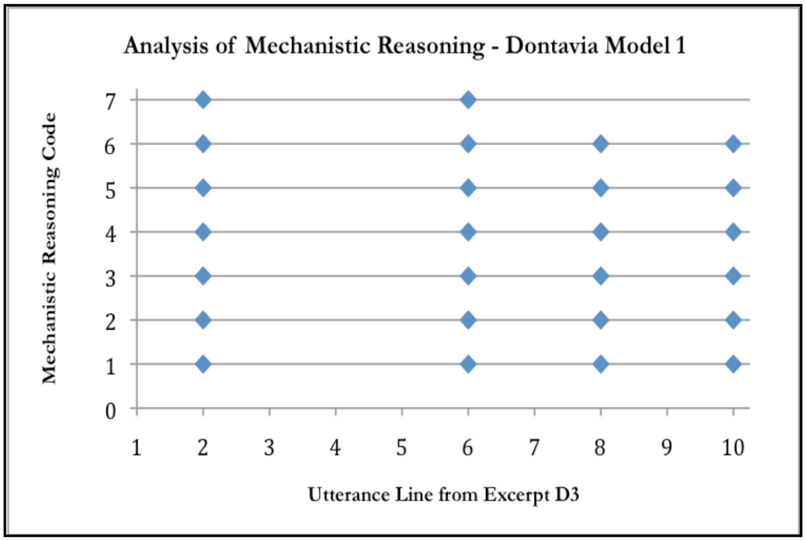


Figure 8: Analysis of Dontavia's Mechanistic Reasoning during her interaction with Model 1

After Dontavia had interacted with Model 2 for approximately half an hour, the lead researcher asked her what the graph in the NetLogo simulation represented. She explained that the graph showed the butterflies' energy. In her explanation, she compared the simulated graph to her own (embodied) experience as a butterfly and the energy graph that she generated during Phases 1 and 2. She identified both similarities and differences between the two graphs (line 1). She commented that the tall peaks on the computer graph mirror a similar peak in her energy graph (line 1, Excerpt 4); however, she also noted that in her graph, she took a strong dive down to five units of energy and then died, having less energy than the relatively stable graph depicted on the computer (lines 2 through 4). Therefore, similar to Model 1, her interaction with Model 2 was also shaped by her previous experience in Phases 1 and 2.

Excerpt 4:

Dontavia: It is the same because my butterfly graph kept going up and down. But it's also different because on this [points to the NetLogo graph], it shows I was up right here and then I went down right here and then I died. But when I died [in the embodied modeling activity], I had only five left. In the end I had less energy than it shows on there [the computer].

Dontavia used an egocentric perspective in order to explain both the graphs. This is evident in her use of the pronoun "I" as she describes the motion of the butterfly (lines 3, 4, and 5). That is, she projected herself as the *agent* or the *entity* in the system (Category 3). She explained that the graphs represented how her energy – i.e., a *property* of the agent (Category 5) - was changing over time. She also identified an *activity* of the agent (Category 4) – death – in order to explain why the graph showed a value of zero. This episode therefore shows that her mechanistic understanding of the mathematical inscription in the simulation, and the change of the butterfly's energy over time represented by that inscription, was directly based on her previous experiences of embodied modeling and graphing.

Classroom Level Findings. Our analysis of students' interviews and written responses in Phase 3 reveals that similar to Dontavia, the other students in the class also productively used their knowledge gained from the embodied modeling activities. We found that 13 students (86%) used backward or forward chaining in order to explain the effect of the new variables that were introduced in Model 1. These students systematically investigated the known variables first, and then were able to investigate the effect of camouflage on the survival of butterflies. This is evident in both their written responses as well as interviews. These responses indicate that like Dontavia, students' earlier embodied actions were an important launching point from which to explore the new model.

In their interactions with Model 2, we found that similar to Dontavia, 14 students (93%) successfully used their previous experience generating discrete mathematical graphs of energy in interpreting the graphs of energy consumption over time displayed in the NetLogo simulation. That is, they were able to identify and explain the similarities between the two graphs. In their written explanations, they successfully identified the entities (agents), their properties and activities, based on which they identified similarities and differences between the simulated graphs and graphs of their energy change during the embodied foraging activity, thereby showing evidence of Levels 3, 4 and 5 of Russ et al.'s framework of mechanistic reasoning. A summary of these findings can be found in Table 4.

Table 4: Summary of Students Strategies when Interaction with the ABMs (N=15)

Model	**Strategy**	**Number of Students**	**Level of Mechanistic Explanation**
Model 1: Camouflage	Known system variables were accounted for by changing parameters to settings that were conducive to survival of the butterflies	13 (86.7%)	7
	Students set parameters based on prior experience, but did not show evidence of testing the effect of new variables through eliminating the effect of known variables.	2 (13.3%)	6
Model 2: Energy	Used energy graphs generated during Phase 2 to explain graphs produced in this model	14 (93.3%)	3,4 & 5
	Uncodeable	1 (6.7%)	N/A

RQ 3: Reasoning about Complexity.

Advantage of Long Proboscis and Functional Equivalence. After completing two iterations of the embodied modeling and representational activities, students came to understand that 1) short flowers provided equivalent energy gains for both long and short proboscised butterflies, and 2) that a long proboscis was biologically advantageous because all food sources in the system were equivalent. To illustrate this conceptual understanding, we provide examples of students reasoning through these elements of the system in Excerpts 5 and 6 below. Excerpt 5 is an example of a single student's thinking concerning the phenomenon of functionally equivalent systems and Excerpt 6 provides a classroom level analysis in the form of a classroom discussion.

Beginning with excerpt 5 below, we see that a student, Jayla, has recorded on her summative worksheet that during Iteration 2, her proboscis could reach nectar from all flowers (lines 1-3). The researcher pushes for her to explain why, during Iteration 2, her proboscis could reach nectar from all flowers to which she responds that she had a long proboscis during Iteration 2 (line 5). In terms of the SBF framework, this excerpt makes explicit that Jayla had recognized the advantage of having a long proboscis (a structure) in terms of its behaviors (drinking from any flower) and function (energy gain).

Excerpt 5:

Researcher	[Reading Jayla's written work aloud] *"The second time I drank nectar my proboscis could reach nectar from all flowers"* Why? Why could your proboscis reach nectar from all flowers?
Jayla	Because it was long.

We found evidence of similar forms reasoning in the classroom discussion that concluded the second iteration of Phases I and II, as shown in Excerpt 6. The researcher begins by asking a student if their proboscis limited them in any way during their forage (line 1). The student responds, similar to Jayla, that their proboscis did not limit them because they had a long proboscis and could drink from any flower (lines 2-3). This provides evidence that like Jayla, the student recognized the functional advantage of having a long proboscis. The researcher then asks the class to expand on that idea (line 7) and students respond that they went to flowers based on the size of their proboscis as well as the location of the flowers (line 8). Following this statement, another student with a long proboscis states that they also went to short flowers even though they had a long proboscis (line 11). The discussion concludes with students suggesting that, overall, the outcome of having a long proboscis is the added advantage of being able to drink from any resource in the system (lines 15 and 17). This suggests that students came to see short and long stemmed flowers as functionally equivalent for long-proboscised butterflies.

Excerpt 6:

1	Researcher	Did your proboscis limit you?
2	Kennedy	No, because I had a long proboscis and I could drink from any
3		flower.
4	Researcher	Ah. Raise your hand if you made a decision to go to flowers
5		because of your proboscis?
6	Class	[All students raise their hands]
7	Researcher	Tell me more about that, Jamar.
8	Jamar	'Cause my proboscis was little and I found closer flowers that
9		were, like, little and short.
10	Researcher	Did you have something to add?
11	Azariah	I went to short flowers but I had a long proboscis.
12	Researcher	Hmm, who had a long proboscis?
13	Class	[Students with a long proboscis raise hands]
14	Researcher	Was a long proboscis an advantage or disadvantage?
15	Julius	An advantage.
16	Researcher	Why?
17	Julius	Because I could go to long or short flowers.

Population Growth and Survival. In this section of the analysis, we discuss how students came to see camouflage as a key mechanism of population survival within the system in

Phase 3. We first present a classroom level analysis of students' written responses which show that a majority of the students were able to identify multiple, loosely coupled factors that would aid the survival of butterflies and population growth. Following this analysis, we present two cases which illustrate the importance of perspective taking when reasoning about inter-agent interactions. The first case highlights the following: a) scaffolds provided by the instructors (including researchers) in response to students' difficulties in reasoning about camouflage, b) how students came to understand camouflage as an *event* that involves interactions between birds, butterflies and flowers and c) evidence that students were able to use camouflage as a mechanism to explain survival and population growth. The second case is a contrasting case which illustrates the importance of being able to adopt the perspective of both the bird and butterfly in understanding camouflage, and highlights how adopting only an egocentric perspective can sometimes lead to erroneous reasoning.

Classroom-Level Analysis of Students' Written Explanations of Survival. In this section, we present an analysis of students' written explanations to the following question: "List all of the things important to a butterfly that wants to survive". Students responded to this question after completing their interactions with Models 1 and 2. During analysis, we found that two students did not fully respond to this question, having left part of their activity sheet blank. In the models designed for this study, reasoning about the survival of butterflies involved reasoning about multiple factors: location of flowers, proboscis length, flower length, butterfly color and flower color. In terms of flower location, butterflies near high-density areas of flowers have a greater chance of survival because they can drink nectar with minimal energy loss during foraging, and therefore have a higher rate of survival than butterflies that are not near an area of high flower density. However, not all butterflies near clumped flowers survive— e.g., an uncamouflaged butterfly will have a greater chance of being eaten by predators. Additionally, not all butterflies can drink from all flowers—having a short proboscis excludes some butterflies from being able to forage from flowers that have nectar deeper than they can reach, no matter how close those flowers are. We found that ten out of fifteen students (67%) identified both proboscis length and flower length ("long proboscis", "proboscis" "flowers" and "drink nectar" on student worksheets) as important factors for survival. Out of these ten students, nine students also identified camouflage ("blending in" and "color" on student worksheets) as an important factor for survival. Thus, a majority of the students (60%) were able to identify camouflage, proboscis length and flower lengths as key factors for the survival of the butterflies.

Note that although this statement asks students to think about *a* butterfly, the goal of the activity was to create conditions in the simulation that would result in a thriving or growing population of butterflies, i.e., an aggregate-level outcome. As explained earlier, we consider this as pedagogical support to foster agent-aggregate complementary reasoning about population growth. Similar to the forms of agent-aggregate reasoning about population growth identified by Dickes & Sengupta (2013) and Wilkerson-Jerde & Wilensky (2014), we therefore believe that our analysis here shows that as a result of participating in this activity, a majority of the students were able to develop a deep understanding of an aggregate-level outcome – population growth of butterflies – in terms of agent-level behaviors, attributes and interactions.

Case 1: Identifying Camouflage as a key interaction for survival and population growth. We begin our analysis by examining two students', Brian and Monikia's, early thoughts on the role of camouflage in terms of the health of the system. Similar to Dontavia, Brian and Monikia had determined through eliminating known variables that color, specifically the color of both the butterflies and flowers, was the most important mechanism of survival (line 1 in Excerpt 7).

When asked by the researcher to explain why this was important in terms of the survival of the butterflies, Brian and Monikia offer explanations from the butterflies' perspective. Specifically, they state that the butterflies get more nectar when they are the same color as the flowers (lines 4 and 6), implying a structure-function relationship between like-colored butterflies and flowers similar to the structure-function relationship between proboscis length and flower length. In lines 10 through 12, the researcher pushes for a change of perspective, asking Brian and Monikia to think like the bird and consider what the bird needs to do to catch a butterfly. Monikia responds that birds need to 'watch' their surroundings (line 14), using their eyes to determine which butterflies are easy to catch. The researcher then guides Brian and Monikia through an egocentric example, selecting Brian as the bird, Monikia as a green butterfly flying among the green grass and themselves as a brown butterfly flying among the green grass (lines 17-19 and 21-23). Without prompting, Brian exclaims that he would 'rather eat [the researcher]' because the researcher is not the same color as the grass (lines 24 and 26), whereas Monikia is the same color as the grass ("she has green and matches the grass", see line 29).

Excerpt 7:

Brian & Monikia	If the butterflies are the same color as the flower they'll stay alive.
Researcher	Why?
Monikia	Because they gain more energy.
Researcher	Why? Who gains more energy?
Brain & Monikia	The butterflies.
Researcher	Why do the butterflies gain more energy if they're the same color as the flower?
Brain & Monikia	We don't know.
Researcher	Okay, so you're thinking like the butterfly now, right? Think like the bird now. So, what does a bird have to do to catch a butterfly?
Brian	Chase it.
Monikia	Watch.
Researcher	Watch. Okay. Watch means…which organ do you use?
Brian & Monikia	Your eyes

17	Researcher	Eyes, right. So you see. Think about this, you're the bird
18		and you're the butterfly. You're wearing green, right?
19		You're out there on the green grass.
20	Brian	I wouldn't rather eat the green butterflies.
21	Researcher	So she's the bird and you're the butterfly. And now
22		imagine me, I'm wearing brown and I'm walking on the
23		green grass.
24	Brian	I'd rather eat you.
25	Researcher	Why?
26	Brian	Because you're not the same color as the grass.
27	Researcher	Not the same color as the grass, right? But why? Why
28		would you eat me and not you?
29	Brian	Because she has on green and she matches the grass, and
30		you have on brown on and you don't match the grass.

After an unsuccessful prompt for Brian and Monikia to explain why the bird would rather eat the non-camouflaged butterfly, the researcher left Brian and Monikia to explore the model further, scaffolded by their activity sheets. The activity sheets were designed to scaffold learners' investigations as they explored the variable space of the model, providing them a space to record their relevant parameters (e.g., values and states of the variables presented in the form of "sliders" in the simulation interface) as well as the aggregate level outcomes of each run of the model (ex. Did your butterflies survive? Is there something you can change to help the butterflies survive?). Approximately fifteen minutes later, the researcher returns and observes Brian and Monikia running a simulation with yellow flowers and yellow butterflies. Once again, the researcher poses the question why birds do not eat butterflies that are same color as the flower (Excerpt 8). Note that Brian first comments that the population of butterflies is growing faster than the birds (lines 2 and 3). The researcher then pushes for Brian to explain this, asking why the birds are not eating butterflies that are the same color as the flowers (lines 4 and 5). Brian explain that the butterflies and flowers 'all look the same' (line 6). Monikia then further explains that both the flowers and butterflies have "black on yellow" (line 7). Thus, Brian and Monikia are now able to identify camouflage as a factor that can explain population growth of the butterflies.

Excerpt 8:

Researcher	What do you think is the reason why [birds]
Brian	[Butterflies] are growing faster than the birds.
Researcher	Oh, that's pretty neat. So, why wouldn't a bird eat the butterflies if they are the same color as the flower?
Brian	Because they (*points to butterflies and flowers*) all look the same.
Monikia	Because they both [flowers and butterflies] have black on yellow.

After the exchange presented in Excerpt 8, Brian and Monikia recorded on their activity sheets that 'blending in' and being the 'same color' as the flowers were important mechanisms in the survival of the butterflies. For Brian and Monikia, an agent-level and egocentric perspective was not a hindrance in reasoning about the stability of the system; however, identifying the appropriate rules of interaction necessitated that Brian and Monikia think like *both* the bird and the butterfly, as well as take into account the relevant attributes of the flowers. With the support of the researcher, Brian and Monikia came to understand camouflage as an event that involves interactions between three types of agents: birds, butterflies and flowers.

*Case 2: The Contrasting Case of Julius and Shalaya***.** The ability to take on the additional perspective of the bird was a key feature in learner's ability to reason about camouflage. Based on our analysis, students who were not able to reason from a different agent-level perspective were unable to successfully identify the role of camouflage in the system. To illustrate this point, we offer the case of Julius and Shalaya. Like their classmates, Julius and Shalaya were able to successfully organize the entities within the system to make foraging decisions that were beneficial to their survival as a butterfly. Although successful during Phase 1 and 2 of the activity sequence, Julius' and Shalaya's overly egocentric view of the system prevented them from successfully identifying the role of camouflage in the simulation.

Excerpt 9:

Researcher	Explain to me what you did today?
Shalaya	The reason our butterflies lived is because we took away the birds. And the reason some of our butterflies died is because we added birds.
Researcher	So, is that the only reason why the butterflies were dying, because the birds were killing them? Was there any other reason the butterflies died?
Shalaya	No.
Researcher	Julius?
Julius	So, the reason we don't have any more birds is because she wants more butterflies in there and I think we should have more birds.
Researcher	Let's add birds. Click on the button, let's add two birds. So you added birds, shouldn't that kill off the butterflies? You added birds, yet the butterflies are living.
Julius	The butterflies are living because they are drinking nectar from each flower so they can get their energy.
Researcher	But why would…in some cases why do the butterflies die out?
Julius	Because they won't be able to move anymore if they die out. If they run out of nectar they won't be able to move anymore.
Researcher	So, the birds have nothing to do with the death of butterflies?
Julius	The birds have nothing to do with the nectar the butterflies drink.
Researcher	So, what happens if we change the color of the butterflies to yellow?
Julius	They're going to die.
Researcher	Why?
Shalaya	That's what happened to ours, look.
Researcher	But why, why would they die if you change the color to yellow? Why didn't they die when you changed the color to red? What's going on? So, now the butterflies are living when we changed the color to red, but when you change it to yellow, they're not. Why do you think the red ones are living and the yellow ones died?
Julius	The red ones drink nectar from each flower and they lived.
Researcher	But why did the yellow ones not live?
Julius	Because that's not the right color of nectar for them to drink.
Researcher	Okay, that could be one reason. What are some other reasons? Think like the bird.
Julius	If they can't get the nectar from each flower they won't be able to live as long.
Researcher	Imagine you're the bird. Okay? I'm a red butterfly and she's a yellow butterfly and this is a red flower, who would you eat first?
Shalaya	Me!

Julius	*(gestures to Shalaya)*
Researcher	Why?
Julius	Because that red flower doesn't compare with the yellow butterfly.
	That yellow butterfly doesn't compare with the red flower.
Researcher	When you say they don't compare, what do you mean?
Julius	They don't, the yellow butterfly is not familiar with that particular
	nectar.
Researcher	What if I told you that all butterflies can drink from any flower nectar.
Julius	They won't be able to.
Researcher	But they all can.
Julius	They're not supposed to drink from a different color flower!

In their exploration of Model 1, during an early investigation, Julius and Shalaya randomly selected to change the color of the butterflies to red at the beginning of their simulation run. Unwittingly, Julius and Shalaya managed to keep their butterflies alive very early on in their investigations, but in the conversation with the researcher (Excerpt 9), despite the color match between flowers and butterflies, they do not identify camouflage or its role in survival. It is clear that they recognize that that the presence of the birds affects the butterfly's survival in some way (lines 3 and 4), going so far as to notice that a lack of birds will cause an increase in the population of the butterflies (line 10). However, they struggle to explain the mechanisms behind this population increase and decrease, later reversing their earlier conclusion by stating that the birds have nothing to do with the survival of the butterflies (lines 8 and 21). The researcher prompts Julius and Shalaya to add birds back into their system (line 12) and asks them why the butterflies are still alive even though there are birds present (line 13). Julius offers an egocentric explanation based on his previous experience as a butterfly, stating that the butterflies are living because they have nectar to drink and are able to 'get their energy' (lines 15 and 16). When pushed by the researcher to explain why the butterflies died in some simulations, Julius again offers an explanation based on his own experience. He states that the butterflies die when they run out of nectar, and that when there is no nectar they are unable to move (lines 18 and 19).

At this point, the researcher requests Julius and Shalaya perform an experiment and asks them what might happen if they change the color of the butterflies to yellow (line 23). Both Shalaya and Julius respond that they will die, stating that they have seen this outcome before (line 25). When asked to explain why, Julius responding that perhaps yellow is not the right color nectar to drink (line 28). At this point, similar to their interaction with Brian and Monikia, the researcher pushes for Julius and Shalaya to take the perspective of the birds (lines 38 and 39). However, unlike Brian and Monikia, Julius does not adopt the perspective of a bird in terms of the difficulty of its vision; he simply states that he would eat the yellow butterfly, and when asked to further explain his answer, he reverts to the perspective of a butterfly, stating that butterflies drink nectar based on familiarity of color (line 46) and are functionally 'not supposed to drink from a different color flower' (line 51).

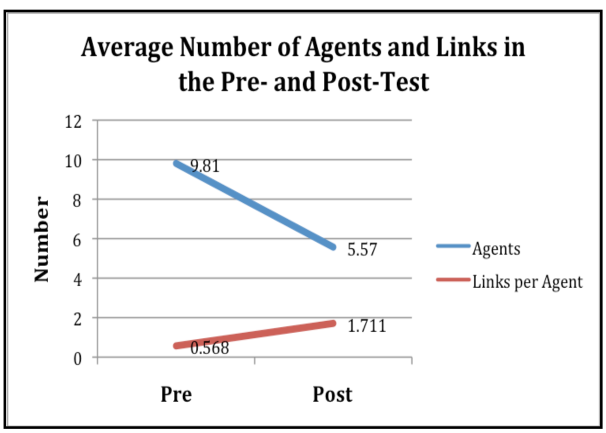
Analytical Summary of Cases 1 & 2. Julius and Shalaya's case provide evidence that simply adopting an egocentric perspective of a typical agent does not entail an appropriate mechanistic understanding of interactions with other agents, especially in the context of understanding camouflage. Julius' refusal to take on the perspective of the bird results in his rejection of the researcher's suggestion that butterflies can drink any type of nectar, irrespective of color. In contrast, Brian and Monikia's case demonstrates that taking on the perspective of *both* the bird and butterflies was essential for developing a mechanistic understanding of camouflage, which in turn helped them explain and understand reasons for butterfly survival. The comparison between these two cases shows that productive strategies for refining learners' understandings of camouflage as a mechanism of survival necessitated taking on the perspectives of *both* birds and butterflies. That is, understanding the factors responsible for survival of an individual agent in an ecosystem requires reasoning not only about the agent itself, but also about its interactions with other (relevant) agents and environmental elements. In terms of mechanistic reasoning, Brian and Monikia's explanation involved reasoning about multiple agents and their interactions (i.e., levels 3, 4 and 5 in Table 1, for both birds and butterflies), whereas, Julius and Shalaya's explanation involved only reasoning about a single agent and its interactions (levels 3, 4 and 5 for only the butterflies).

Pre-Post Comparisons of Representations of Interdependence. Our analysis of student representations of interdependence and energy flow in ecosystems focused on three comparisons: 1) the number of correct links depicted by students (i.e. the flow of energy was drawn from producer to consumer), 2) the number of agents depicted by students and 3) the number of links per agent depicted by students. A summary of our coding scheme for the pre- and post-assessments is provided in the Appendix. Paired sample t-tests were performed for each comparison under an alpha of 0.05. To offset the probability of Type I error due to multiple comparisons on the same data set, we also applied a Bonferroni correction that reduced our alpha from 0.05 to 0.0167. A summary of these findings can be found in Table 5 below.

Table 5: T-test analysis in terms of the change in % correct links of energy flow between the pre- and the post-assessments (N=15, α=0.05, Bonferroni corrected α=0.0167)

Code	Assessment	Outcome	Two-Tailed P-value		T-value	DF
			α=0.05	*α=0.0167*		
Average Correct Links	Pre Post	44.60% 74.87%	<0.05*	<0.05	2.1746	14
Average Number of Agents	Pre Post	9.81 5.57	<0.0013*	<0.0013*	4.0198	14
Average Number of Links per Agent	Pre Post	0.568 1.711	<0.0001*	<0.0001*	6.5030	14

Overall, our analysis of students' representations of energy flow within ecosystems demonstrated that students generated a total of 98 links in the pre-assessment, out of which 36 were correct, i.e., these links depicted normatively accepted direction of energy flow from the producer to the consumer. In the post-assessment students generated a total of 86 links, out of which 65 were correct. That is, across all the students, the total percentage of correct links showing the flow of energy from producer to consumer increased from 36.74% in the pre-assessment to 75.58% in the post-assessment. We also found that each student on average represented 44.60% of their links correctly in the pre-test, and 74.87% of their links correctly in the post-assessment. This gain was found to be statistically significant using a paired sample t-test with an alpha of 0.05 ($p < 0.05$; $t = 2.1746$; $df = 14$), but was not statistically significant after applying a Bonferroni correction (Bonferroni $\alpha=0.0167$). Note that since the probability of Type I Error due to multiple comparisons is only 14%, it is possible that the Bonferroni correction is overly strict, increasing the probability of a Type II Error.



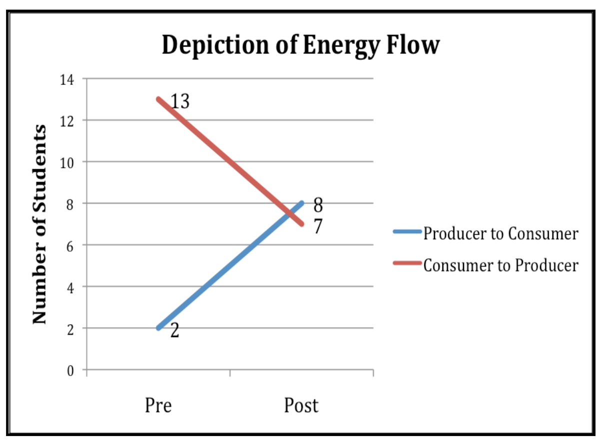


Figure 9: Depiction of Energy Flow in Pre- and Post-tests (N=15)

We also found that students' representations of inter-agent relationships increased significantly in the post-assessment compared to the pre-assessment. We found that the average number of agents students chose to represent decreased in the pre-assessment from 9.81 agents in the pre-assessment to 5.57 agents in the post-assessment. This was evident in terms of the number of links per agent. We found two types of links present in student representations: inter-agent links and agent-environment links. The number of links increased from 0.568 links per agent in the pre-assessment to 1.711 links per agent in the post-assessment. In the pre-assessment, there were many agents with no links; whereas in the post-assessment, on average every agent depicted in the system had at least one link to another agent (either plant or animal) within the system. Further analysis shows that the number of students who depicted energy as flowing from the producer to the consumer increased from 2 in the pre-assessment to 8 in the post-assessment. These findings are shown above in Figure 9.

Another related measure is the ratio of the number of agents depicted by the students and the number of links between agents. Paired sample t-tests reveal that across all students, reductions in the number of agents as well as an increase in the number of links per agent were statistically significant, both with and without the Bonferroni correction. Table 6 summarizes our findings, which suggests that students were able to identify greater number of inter-agent relationships in the post-assessment, compared to the pre-assessment. This finding is consistent with our earlier analysis, which showed that as students progressed through the curricular activities, their' mechanistic explanations progressively increased in sophistication in terms of them being able to identify the various elements within the ecosystem and their relationships.

Discussion

ABMs have been shown to be powerful tools for representing and investigating emergence (Goldstone & Wilensky, 2009). In this paper, our goal was to demonstrate how ABMs could be integrated with 3rd grade elementary science curricula. We began this paper by stating that following Danish et al. (2011), our focus is on the design of activity systems to support students' learning of emergent behaviors using ABMs. Our work highlights two important characteristics of such integration: the importance of embodied modeling in reasoning about interactions in an ecosystem, and the progressive refinement of student reasoning about ecosystems in such an environment.

First, the integration of ABMs in elementary classrooms also necessitates the use of other synergistic forms of modeling. This is a common theme across our work, as well as previous research by Danish and colleagues: inquiry using simulations builds on, and/or complements inquiry activities that children conduct using other forms of modeling such as embodied modeling, as well as generating inscriptions (e.g., drawing and graphs). The path leading to agent-based models of emergent phenomena needs to be *designed* – it cannot simply be assumed that putting a child in front of the simulation may suffice to support her or his inquiry.

But why are these different forms of modeling necessary? Two of these forms of modeling were chosen because prior research shows that one is necessary for the other – i.e., embodied thinking is central to the development of agent-based thinking and representational practices (Papert, 1980; Goldstone & Wilensky, 2009; Wilensky & Reisman, 2006). While it is clear from our analysis that students recalled and built upon their embodied modeling experiences as butterflies during their interactions with the ABMs, it is also important to note that the mathematical inscriptions (bar graphs) also provided a representational continuity between the embodied modeling activities and the ABMs, as well as with previous representational forms that students used and developed in their science and math classes prior to the study. The graphs in the ABMs in Phase 3 were designed specifically to be similar to student-generated graphs during Phase 2. Each student observed the behaviors of several different butterflies by using the "Watch a butterfly” scaffold in the simulation, and provided verbal explanations to both the teacher and the researchers regarding differences among the observed butterflies’ energies and their embodied experiences as butterflies. Furthermore, the order in which nearly all the students explored the variable space in the simulations showed that they first experimented with the variables that they were familiar with during their embodied activity, and then proceeded to explore the newly introduced variables. Thus, our work shows that the integration of embodied modeling and ABMs can productively support students’ inquiry with ABMs.

Although our analysis generally highlights the importance of embodied modeling activities, Julius and Shalaya’s contrasting case also shows that simply adopting an egocentric perspective of the focal agent (in this case, the butterflies) does not always suffice, especially in cases where the focal event involves interactions between different types of agents. Julius and Shalaya’s case shows that reasoning about camouflage by only adopting the perspective of the butterfly can lead to incorrect interpretations of the behaviors and interactions involved – i.e., butterflies only drink nectar from similar color flowers. In contrast, Brian and Monikia were successfully able to adopt the perspectives of both the birds and the butterflies. The researcher specifically prompted them to reason about the vision of the birds, and the difficulties involved (from the perspective of the bird) in identifying camouflaged butterflies from the surrounding environment. That is, in order to develop a deeper understanding of the possible reasons for survival of butterflies (the focal agent), it was necessary to take the perspective of other relevant interacting agents, such as birds, in order to imagine which butterflies would be most visible to the birds. In other words, understanding camouflage required the students to think and reason about the attributes and behaviors of both the birds and the butterflies.

Second, the use of mechanistic reasoning (Russ et al., 2008) as an analytic lens also illustrates the *process* through which students developed a progressively refined understanding of interactions in the ecosystem. We found that students' mechanistic explanations progressively increased in sophistication, in terms of their explanations of the entities of the ecosystem and their interactions, as they iteratively conducted the activities in Phases 1 and 2. According to Russ and colleagues' (2008) coding scheme for mechanistic explanations, students' explanations shifted from Categories 1 - 4 in Phase 1 to "Identification of the Organization of Entities" (Category 6) in Phase II. We see this as a significant conceptual shift in students' understandings. Categories 1 – 4 in the coding scheme focus on identifying agent-level variables and actions, while Category 6 focuses on the organization of these entities and the entailments of their actions. In the following activity (Phase III), students' explanations showed evidence of further refinement as they interacted with the ABMs. With the introduction of new variables in Model 1, student inquiry resulted in *chaining* (i.e., Category 7 in Russ et al., 2008) – i.e., they were able to identify relationships between multiple entities, actions and events. We believe that it was this process of progressive refinement of mechanistic explanations that enabled students to develop more sophisticated representations of inter-relationships within an ecosystem in their post-assessments compared to the pre-assessments.

Modeling is a complex enterprise and science studies tell us that scientific exploration is a long and iterative process that often spans several years from initial model development to eventual acceptance by the scientific community (Latour, 1990; Pickering, 1995). In contrast, the window of children's development we present in this article spans a period of two weeks, as is common in education research. Our study, however, shows that by participating in a set of carefully designed activities that integrated multiple forms of modeling, even in this (comparatively) short period of time, children can indeed begin to develop progressively refined mechanistic explanations of relationships among agents, and between agents and environmental elements within an ecosystem. Furthermore, our analysis also shows that students began to successfully reason about loosely coupled events and functional equivalence, which has been noted to be important characteristics of complex systems in general, and ecological systems in particular. And finally, students were better able to identify correct ecosystem producer-consumer relationships in the post-assessment as compared to the pre-assessment. This comparison also revealed that students were able to identify more relationships between agents in the post-assessment than in the pre-assessment. Note that as we reported earlier, previous studies have argued that students' challenges in learning about food chains and food webs are due to their difficulties in connecting their reasoning about individual agents to more complex interactions and outcomes. We therefore believe that our study shows that the curricular activities reported here can potentially alleviate some of these challenges.

Furthermore, study illustrates an important aspect of pedagogical praxis of modeling complex systems in the science classroom. Danish (2014) points out that in classroom instruction, concepts central to understanding complex systems are often represented in a simplified manner compared to the scientifically normative view (p. 121), and our study is no exception. For example, we have discussed earlier that the state standards and previous instructional history greatly shaped what the teacher considered to be relevant for classroom discussion and productive modeling activities. Her decisions to ask students to develop bar graphs for representing energy change, and her focus on survival of butterflies rather than engaging students in a deeper inquiry around the role of variation in the system are two such examples. We believe that careful studies of such forms of teacher appropriation, rather than focusing predominantly on researcher-led classroom implementations, are crucial for thinking about improvement of complex systems education.

Our work thus adds to the previous research on ABMs in K12 education by Danish and colleagues, and further strengthens the claim that ABMs, which are now a mainstay in the practice of scientists and engineers studying complexity (Chandrasekharan, Nersessian & Subramanian, 2012; Goldstone & Wilensky, 2009), can be made accessible to third graders in the regular science classroom. This work, therefore, suggests a beginning of what might be a long, productive journey in which children can engage in modeling complexity using ABMs for extended periods of time and in a manner that would more closely follow the practices of scientists and engineers.

References

Abrahamson, D., & Wilensky, U. (2004). ProbLab: A computer-supported unit in probability and statistics. In *Proceedings of the 28th annual meeting of the International Group for the Psychology of Mathematics Education* (Vol. 1, p. 369).

Alibali, M. W., & Nathan, M. J. (2012). Embodiment in mathematics teaching and learning: Evidence from learners' and teachers' gestures. *Journal of the Learning Sciences*, *21*(2), 247-286.

Alparslan, C., Tekkaya, C., & Geban, O. (2003). Using the conceptual change instruction to improve learning. Journal of Biological Education, 37(3), 133–137.

Barsalou, L. W. (2008). Grounded cognition. *Annual Review Psychology*, *59*, 617-645.

Basu, S., Sengupta, P., & Biswas, G. (2015). A scaffolding framework to support learning of emergent phenomena using multi-agent based simulation environments. *Research in Science Education*. 45 (2), 293-324.

Basu, S., Biswas, G., & Sengupta, P. (2011). Scaffolding to support learning of ecology in simulation environments. In: Susan Bull and Gautam Biswas (Eds.), *Artificial Intelligence in Education* (pp. 417-419). Springer.

Bechtel, W. and Abrahamsen, A. (2005). Explanation: A Mechanistic Alternative. Studies in History and Philosophy of the Biological and Biomedical Sciences , 36, 421-441.

Bechtel, W., & Richardson, R. C. (1993). Discovering complexity. *Princeton, NJ: Princeton UP*.

Blikstein, P., & Wilensky, U. (2009). An atom is known by the company it keeps: A constructionist learning environment for materials science using multi-agent simulation. *International Journal of Computers for Mathematical Learning*.

Bolger, M. S., Kobiela, M., Weinberg, P. J., & Lehrer, R. (2012). Children's Mechanistic Reasoning. *Cognition and Instruction*, *30*(2), 170-206.

Brandon, Robert (1985), "Green on Mechanism and Reductionism: More Than Just a Side Issue", in Peter Asquith and Philip Kitcher (eds.), PSA 1984, v. 2. East Lansing, MI: Philosophy of Science Association, 345-353.

Brown, M. H., & Schwartz, R. S. (2009). Connecting photosynthesis and cellular respiration: Preservice teachers' conceptions. Journal of Research in Science Teaching, 46(7), 791–812.

Burian, Richard M. (1996), "Underappreciated Pathways Toward Molecular Genetics as Illustrated by Jean Brachet's Cytochemical Embryology", in Sahotra Sarkar (ed.), The Philosophy and History of Molecular Biology: New Perspectives. Dordrecht: Kluwer, 671-85.

Chandrasekharan, S. (2009). Building to discover: a common coding model. *Cognitive Science*, 33(6), 1059-1086.

Chandrasekharan, S., Nersessian, N.J., & Subramanian, V. (2012): Computational Modeling: Is this the end of thought experimenting in science? In J. Brown, M. Frappier, & L. Meynell, eds. *Thought Experiments in Philosophy, Science and the Arts.* (London: Routledge). pp. 239-260.

Chandrasekharan, S., & Nersessian, N. J. (2014). Building Cognition: The Construction of Computational Representations for Scientific Discovery.*Cognitive science*.

Chave, J., Muller-Landau, H.C. & Levin, S.A. (2002) Comparing classical community models; theoretical consequences for patterns of diversity. American Naturalist 15, 1–22.

Chi, M. T. H. (2005). Commonsense Conceptions of Emergent Processes: Why Some Misconceptions Are Robust. *Journal of the Learning Sciences, 14*, 161-199.

Corbin, J. M., & Strauss, A. (1990). Grounded theory research: Procedures, canons, and evaluative criteria. *Qualitative sociology*, *13*(1), 3-21.

Crick, F. (1988), What Mad Pursuit: A Personal View of Scientific Discovery. New York: Basic Books.

Dagher, Z. R., & BouJaoude, S. (1997). Scientific views and religious beliefs of college students: The case of biological evolution. Journal of Research in Science Teaching, 34(5), 429–445.

Danish, J. A. (2014). Applying an activity theory lens to designing instruction for learning about the structure, behavior, and function of a honeybee system. *Journal of the Learning Sciences*, *23*(2), 100-148.

Danish, J. A., & Enyedy, N. (2007). Negotiated representational mediators: How young children decide what to include in their science representations. *Science Education*, *91*(1), 1-35.

Danish, J.A., Peppler, K., Phelps, D., Washington, D. (2011). Life in the Hive: Supporting Inquiry into Complexity Within the Zone of Proximal Development. *Journal of Science, Education and Technology*, DOI 10.1007/s10956-011-9313-4

Dickes, A. C., & Sengupta, P. (2013). Learning natural selection in 4th grade with multi-agent based computational models. *Research in Science Education* 43 (3), 921-953.

Dickes, A. C., Farris, A.V., Sengupta, P., Brutcher, E., & Ghassoul, C. (April, 2015). *Programming motion and ecology as mathematizing: A year long developmental study.* Paper accepted for presentation at the Annual meeting of the American Educational Researchers Association, Chicago, IL.

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

Dreyfus, H. L. (1996). The current relevance of Merleau-Ponty's phenomenology of embodiment. *The Electronic Journal of Analytic Philosophy*,*4*, 1-16.

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

Eilam, B. (2012). System thinking and feeding relations: Learning with a live ecosystem model. Instructional Science, 40(2), 213-239.

Enyedy, N. (2005). Inventing mapping: Creating cultural forms to solve collective problems. *Cognition and Instruction*, *23*(4), 427-466.

Giere, R.N. (1988). *Explaining Science: a cognitive approach*. University of Chicago Press. [RNG]

Glennan, S. (2002). Rethinking mechanistic explanation. Philosophy of science,69(S3), S342-S353.

Goldstone, R. L., & Wilensky, U. (2008). Promoting transfer by grounding complex systems principles. *The Journal of the Learning Sciences*, *17*(4), 465-516.

Gravemeijer, K., Cobb, P., Bowers, J., & Whitenack, J. (2000). Symbolizing, modeling, and instructional design. *Symbolizing and communicating in mathematics classrooms: Perspectives on discourse, tools, and instructional design*, 225-274.

Hammer, D., Russ, R., Mikeska, J., & Scherr, R. (2008). Identifying inquiry and conceptualizing students' abilities. *Teaching scientific inquiry: Recommendations for research and implementation*, 138-156.

Hammer, D., Elby, A., Scherr, R. E., & Redish, E. F. (2005).Resources,framing, and transfer. In J. Mestre (Ed.), Transfer of learning from a modern multidisciplinary perspective (pp.89-120). Greenwich, CT: Information Age Publishing.

Hmelo-Silver, C. E., Eberbach, C., & Jordan, R. (2014). Technology-Supported Inquiry for Learning about Aquatic Ecosystems. *Eurasia Journal of Mathematics, Science & Technology Education*, 10(5), 405-413.

Hmelo-Silver, C. E., Marathe, S., & Liu, L. (2007). Fish swim, rocks sit, and lungs breathe: Expert-novice understanding of complex systems. *The Journal of the Learning Sciences*, *16*(3), 307-331.

Hmelo-Silver, C., & Pfeffer, M. (2004). Comparing expert and novice understanding of a complex system from the perspective of structures, behaviors, and functions. *Cognitive Science, 28*(1), 127-138.

Holland, J. H. (1999). *Emergence: From chaos to order*. Da Capo Press.

Hubbell, S. P. (2005). Neutral theory in community ecology and the hypothesis of functional equivalence. *Functional ecology*, *19*(1), 166-172.

Hubbell, S. P. (2001). *The unified neutral theory of biodiversity and biogeography (MPB-32)* (Vol. 32). Princeton University Press.

Huberman, A. M., & Miles, M. B. (1994). Data management and analysis methods.

Ingold, T. (2013). *Making: Anthropology, archaeology, art and architecture*. Routledge.

Jacobson, M., & Wilensky, U. (2006). Complex systems in education: Scientific and educational importance and implications for the learning sciences. *The Journal of the Learning Sciences, 15 (1),* 11-34.

Kauffman, Stuart A. (1971), "Articulation of Parts Explanation in Biology and the Rational Search for Them", in Roger C. Buck and Robert S. Cohen (eds.), PSA 1970. Dordrecht: Reidel, 257-272.

Keller, E. F. (1983). *A feeling for the organism: the life and work of Barbara McClintock.* W. H. Freeman & Company.

Klopfer, E., & Resnick, M. (2003). Technologies to support the creation of complex systems models: Using StarLogo software with students. Biosystems, 71, 111–123

Klopfer, E., Yoon, S., & Perry, J. (2005). Using palm technology in Participatory Simulations of complex systems: A new take on ubiquitous and accessible mobile computing. *Journal of Science Education and Technology, 14(3)*, 285-298.

Klopfer, E., Yoon, S., & Um, T. (2005). Teaching complex dynamic systems to young students with StarLogo. *Journal of Computers in Mathematics and Science Teaching*, *24*(2), 157-178.

Kuhn, L., & Reiser, B. (2006). Structuring activities to foster argumentative discourse. Paper presented at the annual meeting of the American Educational Research Association, San Francisco, CA.

Latour, B. (1999). Pandora's Hope: *Essays on the reality of science studies*. London: Cambridge University Press.

Lehrer, R., Schauble, L. (2000). Inventing Data Structures for Representational Purposes: Elementary Grade Students' Classification Models. *Mathematical Thinking and Learning*, 2 (1&2), 51-74.

Lehrer, R., & Pritchard, C. (2002). Symbolizing space into being. In *Symbolizing, modeling and tool use in mathematics education* (pp. 59-86). Springer Netherlands.

Lehrer, R., & Schauble, L. (2006a). Scientific thinking and science literacy. In W. Damon, R. Lerner, K.A. Renninger, and I. E. Sigel, Eds. *Handbook of child psychology, 6th Edition, Volume 4: Child psychology in practice* (pp. 153-196). Hoboken, NJ: John Wiley and Sons.

Lehrer, R., & Schauble, L. (2006b). *Cultivating model-based reasoning in science education*. Cambridge University Press.

Lehrer, R., & Schauble, L. (2010). What Kind of Explanation is a Model?. In *Instructional explanations in the disciplines* (pp. 9-22). Springer US.

Lehrer, R., & Schauble, L. (2011). Designing to support long-term growth and development. In *Theories of learning and studies of instructional practice* (pp. 19-38). Springer New York.

Levy, S. T., & Wilensky, U. (2008). Inventing a "mid-level" to make ends meet: Reasoning through the levels of complexity. *Cognition and Instruction, 26(1)*, 1-47.

Liu, X., & Lesniak, K. (2006). Progression in children’s understanding of the matter concept from elementary to high school. Journal of Research in Science Teaching, 43(3), 320–347.

Lin, C. Y., & Hu, R. (2003). Students' understanding of energy flow and matter cycling in the context of the food chain, photosynthesis, and respiration. *Int. J. Sci. Educ.*, *25*(12), 1529-1544.

MacLeod, M. & Nersessian, N.J. (2015): *Modeling Systems-Level Dynamics: Understanding without Mechanistic Explanation in Integrative Systems Biology.* To appear in Studies in History and Philosophy of Science Part C — Biological and Biomedical Science (in press).

Machamer, P., Darden, L., & Craver, C. F. (2000). Thinking about mechanisms.*Philosophy of science*, 1-25.

Mathis, W. J. (2010). The "Common Core" standards initiative: An effective reform tool. *Boulder and Tempe: Education and the Public Interest Center & Education Policy Research Unit. Retrieved July*, *29*, 2010.

Metz, K. (2004). Children's understanding of scientific inquiry: Their conceptualization of uncertainty in investigations of their own design. Cognition and Instruction, 22, 219-290.

Mitchell, M. (2009). *Complexity: A guided tour*. Oxford University Press.

National Research Council. (2008). Taking science to school: Learning and teaching science in grades K-8. Washington, DC: National Academies Press

Nersessian, N. (1992). How do scientists think? Capturing the dynamics of conceptual change in science. In R. Giere (Ed.), Cognitive models of science (pp. 3-44). Minneapolis: University of Minnesota Press.

Nersessian, N. J. (2008) Creating scientific concepts. Cambridge, MA: MIT Press.

Papert, S. (1980). Mindstorms: Children, computers, and powerful ideas. Basic Books, New York, NY.

Peppler, K., Danish, J., Zaitlen, B., Glosson, D., Jacobs, A., & Phelps, D. (2010). BeeSim: Leveraging Wearable Computers in Participatory Simulations with Young Children.

Petri, J. & Niedderer, H. (1998). A learning pathway in high-school level quantum atomic physics *Science Education,* 20, 1075–88.

Pickering, A. (1995). The Mangle of Practice: Time. *Agency and Science, Chicago*.

Purves, D.W. & Pacala, S.W. (2005) Ecological drift in niche-structured communities: neutral pattern does not imply neutral process. Biological Interactions in the Tropics (ed. D. Burslem). Blackwell Publishers, Oxford.

Rapp, D. N., & Sengupta, P. (2012). Models and Modeling in Science Learning. In *Encyclopedia of the Sciences of Learning* (pp. 2320-2322). Springer US.

Reiner, M., & Eilam, B. (2001). Conceptual classroom environment–a system view of learning. International Journal of Science Education, 23(6), 551–568.

Resnick, M. (1994). *Turtles, termites and traffic jams: Explorations in massively parallel microworlds.* Cambridge, MA: MIT Press.

Resnick, M., and Wilensky, U. (1998). Diving into complexity: Developing probabilistic decentralized thinking through role playing activities. *Journal of the Learning Science,* 7: 153–171.

Russ, R. S. (2006). A framework for recognizing mechanistic reasoning in student scientific inquiry.

Russ, R. S., Coffey, J. E., Hammer, D., & Hutchison, P. (2009). Making classroom assessment more accountable to scientific reasoning: A case for attending to mechanistic thinking. *Science Education*, *93*(5), 875-891.

Russ, R., Scherr, R., Hammer, D., & Mikeska, J. (2008). Recognizing mechanistic reasoning in student scientific inquiry: A framework for discourse analysis developed from philosophy of science. *Science Education*, 93, 875-891.

Sengupta, P., Dickes, A.C., Farris, A.V., Karan, A., Martin, D., & Wright, M. (2015). Programming in K12 Science Classrooms. *Communications of the ACM,* 58(11), 37 – 39.

Sengupta, P., & Wilensky, U. (2009). Learning electricity with NIELS: Thinking with electrons and thinking in levels. *International Journal of Computers for Mathematical Learning, 14*(1), 21-50.

Simon, H. A. (1977). The structure of ill-structured problems. In *Models of discovery* (pp. 304-325). Springer Netherlands.

Stevens, R., & Hall, R. (1998). Disciplined perception: Learning to see in technoscience. *Talking mathematics in school: Studies of teaching and learning*, 107-149.

Stroup, W. M., & Wilensky, U. (2014). On the embedded complementarity of agent-based and aggregate reasoning in students' developing understanding of dynamic systems. *Technology, Knowledge and Learning*, *19*(1-2), 19-52.

Taber, K. S. (2008). Exploring Conceptual Integration in Student Thinking: Evidence from a case study. *International Journal of Science Education*, *30 (14)*, 1915 – 1943

Tan, J., & Biswas, G. (2007, March). Simulation-based game learning environments: Building and sustaining a fish tank. In *Digital Game and Intelligent Toy Enhanced Learning, 2007. DIGITEL'07. The First IEEE International Workshop on* (pp. 73-80). IEEE.

Wagh, A., & Wilensky, U. (2014, August). Seeing patterns of change: Supporting student noticing in building models of natural selection. Paper presented at Constructionism and Creativity, Vienna, Austria.

Wilensky, U. (1999a). *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.

Wilensky, U., & Resnick, M. (1999). Thinking in levels: A dynamic systems perspective to making sense of the world. *Journal of Science Education and Technology, 8(1).*

Wilkerson-Jerde, M. H., & Wilensky, U. J. (2014). Patterns, Probabilities, and People: Making Sense of Quantitative Change in Complex Systems. *Journal of the Learning Sciences*.

Wimsatt, W. C. (1976). Reductive explanation: A functional account. In *PSA 1974* (pp. 671-710). Springer Netherlands.

Yin, R.K. (1994). *Case study research: Design and methods (2nd ed.).* Applied Social Research Methods Series, 5. Thousand Oaks: Sage Publications.

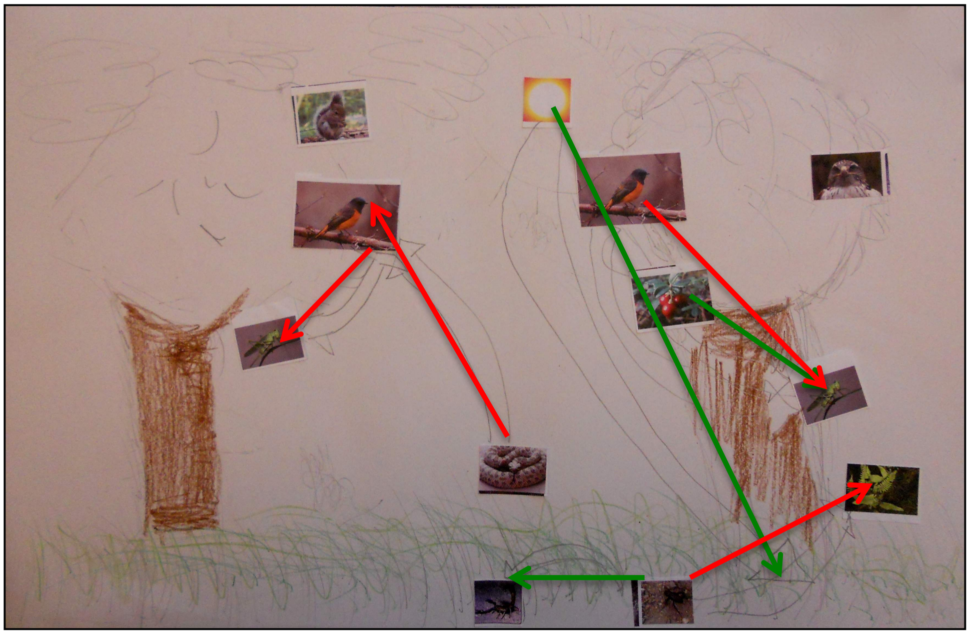
Zamora, R. (2000). Functional equivalence in plant‐animal interactions: ecological and evolutionary consequences. *Oikos*, *88*(2), 442-447.

Appendix

A: Sequence of Learning activities, Duration of each Activity and the Target Learning Goals

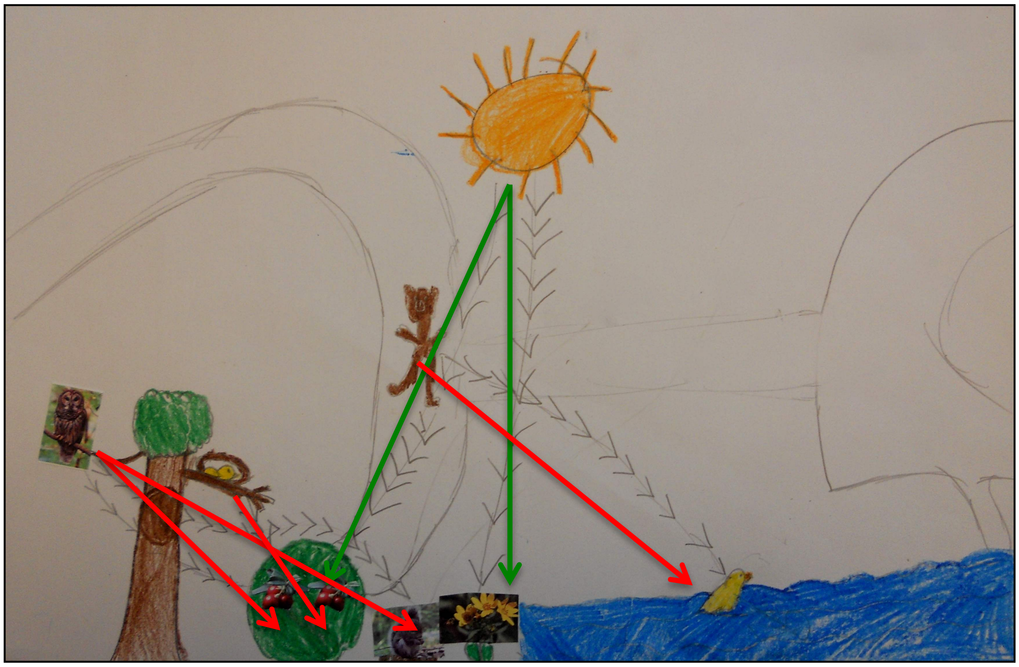
Phase	Day	Activity	Duration of Activities (1hr15min class period)	Target Learning Goals
Pre	**Day 1**	Pre-Test	1hr 15min	Identify relationships between the agents in the system. Identify the flow of energy as moving from the producer to the consumer.
Phases 1 and 2, Iteration 1 and 2	**Day 2**	<u>Iteration 1:</u> Embodied Modeling Bar Graphs & Maps	Embodied Modeling: 25min Energy Maps: 30min Maps: 20min	Introduce students to agent-level rules of the system including: 1) energy gain & losses, 2) Role of Proboscis during forage, 3) Role of flower size during forage and 3) role of flower location during forage Mathematize and spatially orient embodied actions Depict energy change over time
	Day 3	<u>Iteration 2:</u> Embodied Modeling Bar Graphs & Maps	Embodied Modeling: 25min Energy Maps: 30min Maps: 20min	Same as Day 2 described above except, varying proboscis length and forage start point used to further highlight the relationships between the entities of the system.
	Day 4	Summative Review Worksheets & Discussion	1hr 15min	Identify relationships between the butterflies and flowers in the system Identify important mechanisms for butterfly survival
Phase 3	**Day 5**	Computational Model 1 & Activity Sheet	1hr 15min	Camouflage as mechanism for survival and population increase.
	Day 6	Computational Model 2 & Activity Sheet	1hr 15min	Interpretation of contiguous graph of change in energy over time
Post	**Day 7**	Post-Test	1hr 15min	Identify relationships between the agents in the system. Identify the flow of energy as moving from the producer to the consumer.

B: A sample pre-assessment coded for 1) total number of agents, 2) total number of links 3) ratio of links to agents and 4) the percentage of links indicating the flow of energy from producer to consumer.



Analytic Note: In this sample pre-test, there are 15 agents depicted: A1 a tree, A2 a squirrel, A3 a bird, A4 a grasshopper, A5 the sun, A6 a bird, A7 berries, A8 a tree, A9 a hawk, A10 a snake, A11 a scorpion, A12 a spider, A13 a grasshopper, A14 a plant and A15 grass. There are also seven links depicted: L1 between the bird and the grasshopper, L2 between the snake and the bird, L3 between the sun and the grass, L4 between the bird and the grasshopper, L5 between the berries and the grasshopper, L6 between the spider and the scorpion and L7 between the spider and the plant. All links but L4 have the flow of energy depicted with arrows. The ratio of links per agent is 0.47. Of the seven links depicted in the picture, three of them are correct producer to consumer links: from the sun to the grass, from the berries to the grasshopper and from the spider to the scorpion. The remaining three links are incorrect since they show the flow of energy as moving from consumer to producer. The percentage of correct links is 3 out of 7 or 42.9%.

C: A sample post-assessment coded for 1) total number of agents, 2) total number of links 3) ratio of links to agents and 4) the percentage of links indicating the flow of energy from producer to consumer.



Analytic Note: In this sample post-test, there are 9 agents depicted: A1 an owl, A2 a tree, A3 a bird in a nest, A4 a berry bush, A5 a squirrel, A6 daisies, A7 a duck, A8 a bear and A9 the sun. There are also six links depicted: L1 between the owl and the squirrel, L2 between the owl and the berry bush, L3 between the bird and the berry bush, L4 between the sun and the berry bush, L5 between the sun and the daisies and L6 between the bear and the duck. All 6 links have the flow of energy depicted with arrows. The ratio of links per agent is 6 over 9 or 0.67. Of the six links depicted in the picture, two of them are correct producer to consumer links: from the sun to the berry bush and from the sun to the daisies. The remaining four links are incorrect since they show the flow of energy as moving from consumer to producer. The percentage of correct links is 2 out of 6 or 33.3%.

CHAPTER 4

SOCIOMATHEMATICAL NORMS FOR INTEGRATING COMPUTATIONAL THINKING AND MODELING WITH ELEMENTARY SCIENCE[3]

Introduction

Modeling is the language of science (Lehrer, 2009; NGSS, 2013). The integration of modeling and other epistemic practices has been recognized as a central objective for K-12 science education (NGSS, 2013). Over the past few years, the integration of computational modeling in K-12 classrooms has become an important focus of research (Sengupta, Kinnebrew, Basu, Biswas & Clark, 2013; Wilensky, Brady & Horn, 2014; Dickes, Sengupta, Farris & Basu, 2016).

The particular form (genre) of computational programming and modeling we focus on in this paper is agent-based modeling and programming (ABM). The term “agent” in ABM indicates an individual computational object or actor (e.g., a Logo Turtle), which carries out actions based on simple rules that are body-syntonic and therefore intuitive for learners (e.g., moving forward, changing directions, speeding up, etc.). Consequently, it is no surprise that researchers have been arguing for teaching and learning motion in elementary classrooms using agent-based programming since the 1980s (Papert, 1980). Many contemporary ABM platforms employ visual programming interfaces, which makes it even easier for learners to assign or control these rules (Sengupta et al., 2015). In the context of learning kinematics, previous research shows that given appropriate teacher-led scaffolding, middle and high school students

[3] This chapter is under review in *Research and Practice in Technology Enhanced Learning*

can effectively use Logo-based platforms to develop deep understandings and mathematical representations of motion (diSessa, Hammer, Sherin, & Kolpakowski, 1991; Sherin, diSessa, & Hammer, 1993; Sengupta & Farris, 2014). However, the same literature also highlights challenges to classroom adoption of such pedagogical approaches. The high overhead associated with teaching Logo programming and teaching physics can lead to potentially "probibitive" demands on the teacher (Sherin, diSessa, & Hammer, 1993, p. 116). A central challenge stems from the sequestered nature of teaching and learning programming on one hand, and teaching and learning physics using programming on the other, typically requiring a different teacher for each part (Sherin, diSessa, & Hammer, 1993). Our goal is to address this challenge by integrating, and not sequestering, these two forms of instruction.

In this paper, we advance an argument that emphasizing *mathematizing* and *measurement* as key forms of learning activities, through the development of *sociomathematical norms* (McClain & Cobb, 2001; Yackel & Cobb, 1996; Cobb, Wood, Yackel, & McNeal, 1992) can help teachers meaningfully integrate programming as the "language" of science. We report a study in which a third grade teacher, in partnership with researchers, integrated agent-based programming with her regular science curriculum by iteratively developing sociomathematical norms for modeling motion using agent-based computational models.

Theoretical Framework

Computational Modeling and Programming as Science Literacy

Vee (2013) defines "literacy" as "facility with a symbolic and infrastructural technology which can be used for creative, communicative and rhetorical purposes". This definition of literacy borrows heavily from diSessa's (2000) definition of literacy which states that "literacy is a socially widespread patterned deployment of skills and capabilities in a context of material support (that is, an exercise of material intelligence) to achieve intellectual ends". Both scholars note that facility with a material intelligence is an important component of literacy, and both argue that for a material intelligence to become a literacy it must first become "infrastructural" to a society's communicative practices (Vee, 2013; diSessa, 2000), that is there must be widespread ability to compose and interpret with that technology. diSessa argued that literacy of any form, and therefore, computational literacy, involves an interplay between material, social and cognitive dimensions.

Our agenda is to argue for considering computational modeling and programming as an integral component of scientific work in the K12 classroom, and therefore, scientific literacy. The distinction between focusing on coding or programming as isolated competencies and our approach can be understood in light of diSessa's distinction between "material intelligence" and literacies. diSessa (2001) argued that while material intelligence can be understood as meaningful use of a technology, literacies are needed for negotiating their lived worlds. Indeed, computational modeling and programming is indeed a genre of modeling that also involves developing material intelligence - competence with programming languages and modeling platforms - in order to be able to use it for designing scientific models. But our goal is show that computational modeling and programming can cease to exist merely as material intelligence and become a core component of scientific literacy in the K12 classroom, especially when teachers organize instruction in particular ways.

It is now widely agreed upon that modeling is the language of science (Giere, 1988; Nercessian, 1996; Lehrer, 2009; NGSS, 2014), and therefore, learning science must involve learning *modeling* as a key scientific practice. In light of diSessa's three pillars, in the classroom, this involves complex interplay and negotiations between computational and non-computational representations (material), reasoning and discourse (cognitive and social), and the emergent classroom micro-culture (social). Our focus in this paper is to study the complex interplay and negotiations involved *from the perspective of a teacher* with no prior background in programming or computational modeling.

Sociomathematical Norms for Integrating Programming with K-12 Science

Our previous work has demonstrated that bringing about the integration of programming and K-12 science education requires careful attention to the design of programming languages, as well as activity systems. Along the first dimension, we have argued that programming languages should employ both domain-specific and domain-general programming commands (Sengupta, & Farris, 2012; Sengupta et al., 2013; Farris & Sengupta, 2014). Along the second dimension, we have argued that the design of learning activities should seek to tightly couple programming and science. For example, Sengupta and Farris (2012) and Sengupta et al. (2013) proposed an activity sequence in which initial activities can foster necessary competencies, such as thinking like an agent through embodied modeling, which can also help children become proficient with programming syntax, commands and control flow, and practices such as debugging, through activities such as "drawing" simple geometric shapes with their bodies and then modeling the shapes using programming. In later activities, children can use these shapes as models of motion. These studies have shown that as students progress through these activities, they begin to become more fluent in modeling motion as a process of continuous change, which has been shown to be a key conceptual challenge for K16 students (Dykstra & Sweet, 2009).

However, research on integrating programing with the K-12 science curriculum has been largely interventionist in nature (diSessa et al., 1991; Sengupta et al., 2013; Wilkerson-Jerde, Wagh & Wilensky, 2015). In contrast, our work here takes an *integrative* stance, where our role as researchers was largely limited to designing activities in partnership with the teacher, based on what the teacher wanted to accomplish on a day to day basis, as mandated by the state and national science and math standards. We believe that such forms of researcher-teacher partnership, where teachers exercise significant agency in the direction and co-design of the curricular activities and lead the classroom teaching and implementation, are methodologically crucial for addressing the issue of effectively managing the tradeoff between teaching programming and teaching science.

In this paper, we propose that emphasizing mathematizing and measurement as key forms of learning activities can help teachers meaningfully integrate programming as a "language" of science, and further, that teachers can accomplish this by supporting the development of sociomathematical norms. The iterative design of mathematical measures can result in deep conceptual growth of students in elementary science, especially when these activities are integrated throughout the curriculum over several months (Lehrer, 2009). Furthermore, the development of children's scientific and mathematical modeling in the classroom in an authentic manner should also involve and can be greatly benefitted by the iterative development and refinement of collective, (i.e., classroom-level), *normative* modeling practices (McClain & Cobb, 2001; Lehrer, Schauble & Lucas, 2008). Science educators have also shown that the question of what counts as a "good" model also needs to be normatively established in classroom instruction in order to deepen students' engagement with scientific modeling in elementary grades, and that these norms also follow similar shifts toward deeper disciplinary warrants over time (Lehrer & Schauble, 2006; Ford & Forman, 2006; Lehrer, Lucas & Schauble, 2008). Lehrer and colleagues (2000, 2001, 2006, 2008) demonstrated that authentic epistemic work in the science classroom must develop and deepen through the social construction of scientific knowledge, and also highlight mathematics as a meaning-making lens through which the natural world can be systematized and described (Lehrer, Schauble, Strom & Pligge, 2001). The authors argue that an emphasis on measurement, including aspects of measurement such as precision and error, and normatively guided model refinement help students move beyond a focus on superficial features of the target phenomena to modeling "unseen" relationships between variables and underlying mechanisms. This supports a shift from "looks like" to "works like" in children's model development (Lehrer & Schauble, 2000, 2006, 2008).

The specific genre of norms we focus on in this paper have been termed *sociomathematical norms* (McClain & Cobb, 2001; Yackel & Cobb, 1996; Cobb, Wood, Yackel, & McNeal, 1992). Sociomathematical norms differ from general social norms that constitute classroom participation structure in that they concern the normative aspects of classroom actions and interactions that are specifically mathematical. These norms regulate classroom discourse and influence the learning opportunities that arise for both the students and the teacher. In the work of Cobb and his colleagues, teachers initiate and guide the development of social norms in mathematics classrooms that sustain classroom micro-cultures characterized by explanation justification, and argumentation (Cobb, Yackel, & Wood, 1989; Yackel, Cobb, & Wood, 1991). Similar to their work, our focus is on the perspective of the teacher, who initiated these norms on her own accord, without any prompting by the researchers.

An important, and rather fundamental sociomathematical norm is what counts as an *acceptable* mathematical solution, and further, this norm typically originates as a *socially* defined norm, and shifts over time to a *sociomathematically* defined norm (Yackel & Cobb, 1996). However, given that these norms are often teacher-initiated, it is also important to look at how these ideas and opportunities are taken up by students in their work (Gresalfi & Cobb, 2009). Our goal is to demonstrate how the emphasis on developing and refining sociomathematical norms pertaining to the design of mathematical measures of motion can help teachers seamlessly integrate programming with science education in a 3rd grade classroom, and how they are taken up in students' work.

Research Questions

To this end, we investigate the following research questions:

1. What were the sociomathematical norms that developed, and how were they taken up by the students?

2. Did these norms shape in any way the development of students' computational models and computational thinking? If so, how?

Method

The Programming Environment

We used ViMAP (Sengupta, Dickes, Farris, Karan, Martin & Wright, 2015), an agent-based, visual programming language that uses the NetLogo modeling platform as its simulation engine (Wilensky, 1999). In ViMAP (Figure 1), users construct programs using a drag-and-drop interface to control the behaviors of one or more computational agents. ViMAP programming primitives include domain-specific and domain- general commands as well as a "grapher" which allows users to design mathematical measures based on periodic measurements of agent-specific and aggregate-level variables (e.g., speed and number of agents, etc.)

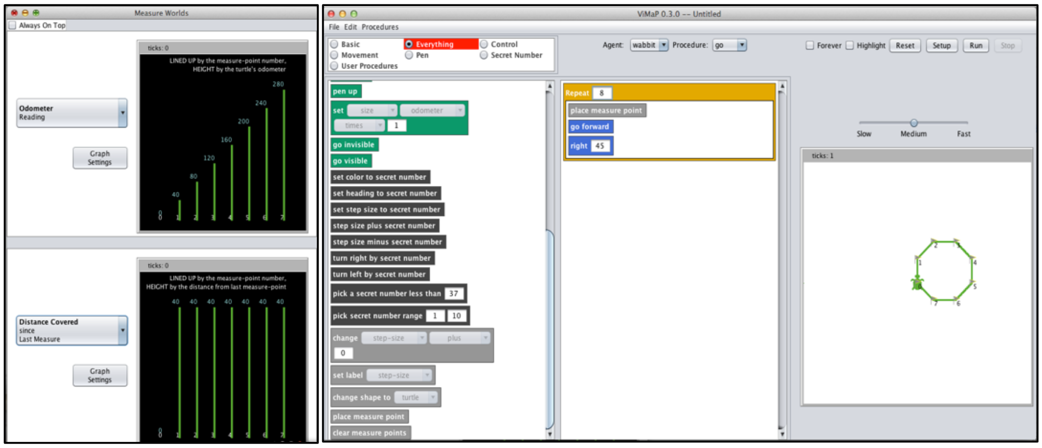


Figure 1: ViMAP's measurement window and programming interface. Figure illustrates the program for generating a regular octagon, the enactment by the turtle agent and graphical representations of length of each line segment (graph on lower left) and perimeter (graph on top left)

Setting & Participants

This study was conducted over the course of 7 months in a 3rd grade classroom in a 99% African-American public charter school located in a large metropolitan school district in the southeastern United States. Fifteen students – fourteen African-American and one Latino – participated in this study. Researchers met weekly with the classroom teacher (Emma, pseudonym) and iteratively co-designed the classroom activities. The teacher taught all lessons, and changes to the activities were made based on her formal and informal assessments of student understanding of the material or in-the-moment responses to student ideas. These adjustments often took the form of extending instructional time on a topic, and modifying the designed classroom materials to better meet the mandated instructional goals. Throughout the year, the teacher emphasized connecting modeling in ViMAP to other out-of-computer modeling experiences, such as embodied and physical modeling activities, as well as re-framed the computational representations in ViMAP as analogous to meaningful lived experiences for both herself and the students. The emphasis on developing classroom-wide conventions was a practice that the teacher employed in her regular math instruction. Our study focuses on how the teacher adapted and employed this approach as a way to integrate modeling motion using ViMAP with her science curriculum.

The learning activities were divided into three phases: Phase I (Geometry), Phase II (Kinematics) and Phase III (Ecology). The present paper reports on Phase II, Kinematics, and traces the development of three normative, mathematical criteria for "what counts" as good ViMAP models of motion. Instruction during Phase II focused on the invention and interpretation of mathematical measures and using ViMAP as a way to explain a real-life phenomenon involving motion (e.g., walking at a constant rate or two cars traveling at different rates for different periods of time). Table 1 summarizes the learning activities during Phase II.

Table 1: Summary of learning activities during Phase II

Activity	Description
Leaving Footprints	Students leave ink footprints on banner paper.
Generating Measures	Students iteratively develop, apply, test and refine a measurement of distance termed a 'step size'.
Collecting step-size data	Students use the 'step size' measurement convention to measure their personal step sizes.
Modeling Step-sizes in ViMAP	Students model and refine their "step-sizes" in ViMAP, generate and discuss "total-distance" graphs & make predictions using ViMAP's grapher.
Modeling Motion as a Process of Continuous Change	Students model motion scenarios in ViMAP and check the validity of those models using ViMAP's grapher and the total distance equation.

Data & Analysis

Data for this work comes from informal interviews with the participants, video recordings of class activities and discussion, student artifacts (e.g. student representations, activity sheets, ViMAP models and pre-, mid, and post-assessments) and daily field notes. The lead researcher and the classroom teacher conducted informal interviews during opportune moments while the students were engaged in single, pair or small group work around modeling and representational activities. Classes were video recorded, and student-created artifacts (ViMAP models, written work) were also collected.

We present the analysis of in the form of explanatory case studies (Yin, 1994), which are well suited as a methodology to answer *how* and *why* questions. We find this to be good fit because our goal here is to illustrate the *process* through which the classroom developed sociomathematical norms, which includes answering *how* the development of these norms shaped the students' interactions with ViMAP and other modeling experiences, and *why* these norms were deemed useful by the teacher. Following previous studies (Dickes et al., 2015), our selection and analysis of cases were guided by the following two criteria: representativeness and typicality.

Representativeness implies that the selected cases should aptly represent key aspects of learning experienced by the students. These key aspects or themes, in turn, are defined based on the research questions. For our purposes, representativeness implies that each case should highlight an important aspect of the process through which the relevant sociomathematical norm emerged. *Typicality* implies that the selected case(s) should potentially offer insights that are likely to have wider relevance for the remainder of the participants in the study. In other words, the cases selected should represent aspects of the process of learning experienced by a majority of the student population. Each of the cases we presented here was typical of majority of the students in the classroom, as evident in comparisons of student work across all students. To this end, as appropriate, we report, a) excerpts from classroom discussion where both consensus and dissent are evident, with a focus on class discussions where deviances from the norm were addressed; and b) an overall (classroom-wide) analysis of take-up of the norms in students' representational work. In addition, to answer RQ2, we also present a classroom-level analysis of students' ViMAP code and models in terms of the quality of their code. We explain the coding scheme later, along with the Findings, and also explain why we believe these changes in students' computational work were shaped by their take-up of the sociomathematical norms.

Findings

The analysis presented below illustrates the development of sociomathematical norms for measuring (Inventing Measures), describing (Approximations) and extending data (Predictions). We explain these norms, describe how they were taken up in student work and, finally, how the development of each norm paralleled new computational practices.

Inventing Measures: Movement from Social to Sociomathematical

Instruction during Phase II began with an investigation of animal tracks. Using the richly illustrated children's book "*Wild Tracks! A Guide to Nature's Footprints"*, Emma and students discussed how animal footprints were data-laden. Among the ideas offered by students and privileged by Emma were animal tracks as histories of "where [the animal] started [moving] and where [the animal] stopped" and whether or not the animal was "running or walking". Each of these ideas emphasize footprints as *measureable objects* and a source of knowledge regarding the behavior of the footprint-leaver. In particular, they leverage footprints as sources of data on the rate and distance traveled by an agent, concepts that frame all student investigations during Phase II. Following this discussion, students generated an embodied artifact, their own footprints inked onto a strip of banner paper and, guided by Emma, problematized the idea of a "step-size". What is a step-size and if we were to measure one, where would we begin measuring and where would we end? Students offered three options for 'step size' measures. Step sizes are measured from heel-to-toe, measured from heel-to-heel and finally measured from toe-to-toe (Figure 2). At this stage, selection of the “best” step-size measurement convention was primarily a social endeavor, with students defining the best step-size measure based on a majority (> 50%) class vote, ultimately selecting the heel-to-toe measurement convention because it was quote, “the biggest” or because their “friend voted for it”, indicating the “socially” grounded nature of what counts as a good measure. Emma encouraged this initial social negotiation, remarking to the class that ultimately “[they] had to make the decision” and that there was no “right or wrong” convention provided implementation of the measure was “consistent” across the class.

Figure 2: Student ideas on how to measure a “step-size”

Following selection of the heel-to-toe step-size measurement convention, Emma instructed students to return to their footprint artifact and measure their unique step-sizes using the heel-to-toe measurement convention. She then asked for students to add up each of those individual step sizes to generate a total distance traveled. Finally, Emma asked the class to measure, with yardsticks or measuring tape, their total distances traveled on their footprint artifacts and record that value on the same data sheet. In a conversation with the researchers while students were engaged in the activity of measuring their unique step-sizes, Emma explained why she wanted students to generate two measures of total distance. During initial negotiation of the “best” step-size measurement convention, Emma recognized the convention the students had selected was not, to use her words, an “accurate” measure of total distance traveled because it produced an overlap, effectively measuring portions of the distance twice (Figure 3). She wanted to encourage students to qualify measurement tools based on their ability to perform the function they are intended to perform, and designed the activities described above to disrupt the class' notion of a "good" step-size measure. Upon completing the activity, students publically reported their findings and found that the measured (ruler on footprint artifact) and the calculated (adding step sizes measured using convention) "didn't match", when they had predicted that they would (Figure 3).

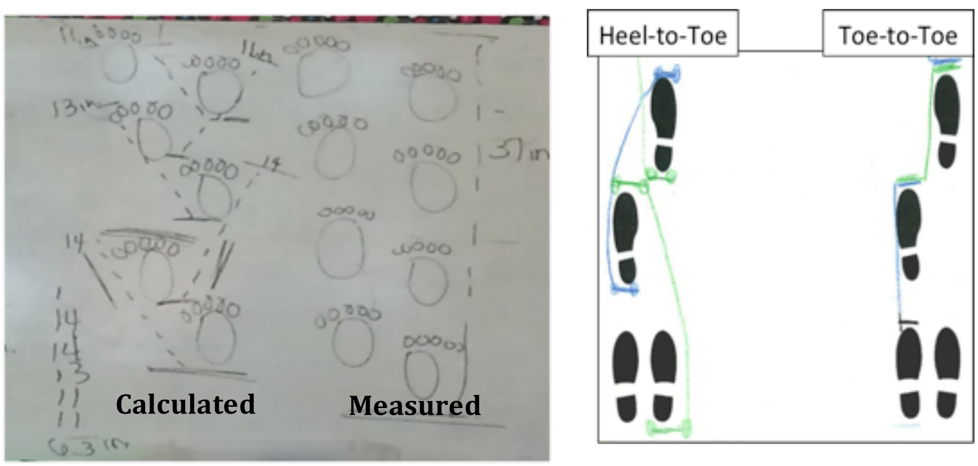


Figure 3: Refinement of "step-size" measurement convention from socially defined (heel-to-toe) to sociomathematically defined (toe-to-toe).

As Emma had predicted, the measured and calculated total distances did not match *due to the inaccuracy of heel-to-toe convention as a measurement of total distance*. This discrepancy prompted Emma to suggest to the class that "maybe we need to find a measure that is more mathematically accurate". Here, "accurate" indicates the ability of the measure to actually measure what the measurer intends to measure, an idea Emma privileged during revision of the step-size measurement convention. Addressing the class, she asked them to explain to her why knowing your step-size was “useful”. Students in the class suggested that step-sizes help you “know how far you went”, indicating step-sizes as measures of distance traveled. However, students continued to disagree on where a step-size should begin. Two students offered competing definitions: 1) measurement begins at the heel of the first foot and 2) measurement begins at the toe of the first foot. Reminding the class to “keep in mind” the question they were trying to answer (“how far someone went”), Emma asks the class to explain to her which one the two “competing schools of thought” they “liked” and *why* they liked them. One student, Marvin, explains to the class that he felt measurement should begin at the toe because “at the toe you *begin to go forward*”. Emma agrees with Marvin’s assessment that distance traveled begins with forward movement, and by validating his idea she provides an opportunity for his idea to be adopted by the remainder of the class. The outcome of this disruption and the discussion that followed was the invention of the “mathematically accurate” toe-to-toe measurement convention, shown in Figure 3.

What is notable in this episode is the development of criteria for what counted as a “good measure”. Initially, “good” measures were socially defined, with students selecting how to measure a “step-size” for reasons unrelated to the purpose of the measure. Following the failure of the heel-to-toe convention, the criteria for “good” measures shifted towards mathematical accuracy. In other words, the value of the measure was assessed on its ability to accurately measure what the measurer intended for it to measure. Emma played an important role in this shift by orchestrating opportunities for normative social definitions of measures to be placed at odds with student activity. At the beginning of students’ exploration, normative criteria for what counted as a “good” step-size measure was neither dependent on the question being asked nor the utility of the measure. In her instruction, the teacher made explicit attempts to highlight the epistemic value of measures, connecting measurement to as a means to answer questions that the students found important to answer ("how far you went") and placing high social value on measures that accurately represented aspects of the phenomena under study (motion, i.e. "go[ing] forward"). Each of these instructional moves was important in deepening student engagement with the phenomena and scaffolding their interactions with modeling motion in ViMAP, as discussed in the sections that follow.

Approximation & Prediction: Norms for Model Refinement

Following the invention of the "mathematically accurate" toe-to-toe measurement convention, students also explored ideas of approximation and prediction as methods for iterating the toe-to-toe measurement convention to quickly find distances that "could not be walked" and refining their ViMAP models of motion phenomena. After students re-measured their step sizes based on the new toe-to-toe measurement convention, the class was asked to think about what value best represented their individual step size data. The teacher introduced 'approximations' as representative values because, as she explained to the researchers, she believed that averages would be difficult for her students. She framed thinking around approximate values as a measurement problem, asking the class to imagine they "wanted to know how far [someone] traveled after fifty or one hundred steps" but "didn't want to actually walk and add up all those different step sizes". Students in the class initially struggled, attempting to solve the problem by embodying the necessary number of steps, and were unable to think about the problem outside of the context of physically measuring and adding steps. Addressing the researchers, Emma explained that she needed to "think of a way" to frame the question "in words [the students] will understand". Refocusing the class, she asks a student, Damien, to read out his step-size data as she physically walks each step in front of the class. She then asks the class what they notice. Damien points out that each step is “changing”. Emma agrees, but follows up with another question: Is each step changing “by a lot” or “by a little”? Instructing the class to “look at [their] data sheets” and examine their own steps, she restates the question. Damien states that his steps “mostly change by a little” and another student, Jayla, agrees stating that she “walked mostly the same” when she measured her individual step sizes. Emma legitimatizes Damien and Jayla’s claims, asking the class to think about what step-sizes that “change by a lot” would “look like”. Embodying step sizes of varying sizes with each step, the teacher asks the students if, when they walk, they “walk like this” or if they “see anyone walk like that down the [school’s] hallway”. The class laughs at the randomness of the teacher’s movements, but agrees with Damien and Jayla’s noticing that individuals take steps of “about the same size” when they walk and that unknown steps would be “close to” the same size as “known” steps.

The teacher used this thought experiment to development normative ideas for “approximate” step-size values. She introduced the term “approximate” on her own accord in order to bring to the children’s attention to the importance of consistency of measured values. Emma provided students with a hypothetical data set, step sizes of 11, 9, 11 and 12, and asked them to build a ViMAP model of the total distance traveled based on the general pattern of the step sizes. To facilitate this, she asked students to reason about the following: if the hypothetical student “continued walking”, what would “their next step be”? In a flurry of discussion, each of the fifteen students offered their ideas. Fourteen of fifteen students (93%) agreed that a “good” possible “future step” was any value already within the range of the set of empirical data, i.e. a value of 9, 10, 11, or 12, suggesting that “good” approximate step sizes were values that were “close to the actual, but not exact” and at the same time, represented the general trend of the values. One student in particular offered to the class that “11” was the best choice because it appeared “the most times” and was “in the middle” of the data set. Only one student deviated from the other students, suggesting that “13” was the next possible step since it “continued the pattern” established by the final two steps of 11 and 12. This deviance from the emerging norm of a “good” approximation was addressed as follows. The teacher referred back to the shared classroom definition of approximation, *close to actual but not exact,* and modeled in ViMAP an approximate step size of ‘13’. She then asked the class to consider the total distance traveled in each model: 43 using actual step size values and 53 using an approximate value of 13. A student responds that the approximate total distance is “too far away from the number [the person] actually walked”. The rest of the class agrees that the two distances are not ‘close’ and came to a consensus that ‘good’ approximate values were “close to” the actual value in terms of both individual step sizes *and* total distance traveled.

Reasoning with approximate values also gave students predictive power through extending their ViMAP models of motion and using multiplicative reasoning. During a teacher-led class discussion on calculating approximate total distances, students noticed that you could use repeated addition (8+8+8+8+8+8) or multiplication (8 x 6) to quickly solve for total distance traveled using approximate steps sizes. The teacher asked what “the formula” for finding total distance would be if they were not “using numbers”. Two students in the class responded that they are multiplying the “number of steps” by the “approximate step size”, generating a formula for total distance: *Total Distance = Number of Steps x Approximate Step Size*. In the teacher’s words, this formula would allow the students to “find total distances that you can’t actually walk”, and therefore, can be used to make predictions.

How did students take this up in their work? An illustrative case is shown in Table 2. In this excerpt, Angelo (a student) interprets the formula as a means to both “win a bet” as well as mathematically verify the accuracy of his ViMAP model of distance. Angelo comments in lines 4 and 5 that if someone bet him that he would only travel less than or equal to 100 units of distance, he would know that they were wrong based on his understanding of approximate values and their role in the total distance formula the class had derived. The researcher affirms Angelo’s observation, asking him if he could prove an acquaintance wrong if he knew his approximate step size (lines 6, 7, 8 and 9). Angelo responds in line 10 that he could. When asked by the researcher how he could prove them wrong (line 11), he offers two possible solutions: the graphs he had generated in his ViMAP model (shown in Figure 4) and his formula (line 14).

Epistemologically, this is a significant move. As Angelo put it, using approximate values allows him to “know” (line 4). We believe that Angelo’s explanation of “betting” and “knowing” here is his intuitive way of explaining what prediction is. Furthermore, this demonstrates that Angelo is able to mathematically summarize discrete values to model continuous patterns of change.

Excerpt 1:

Researcher	How far did you walk after taking 15 steps?
Angelo	300 distance
Researcher	That’s exactly right.
Angelo	So, if somebody bet that I won’t make it farther than 100 *I know* that I will make it.
Researcher	That’s right. That’s how a formula for approximate distance can help you. If someone said “I bet Angelo would only walk 150 inches in 15 steps”, but you knew what your approximate step size was, could you prove them wrong?
Angelo	Yes
Researcher	How?
Angelo	I could look at my graph.
Researcher	Or you could do what?
Angelo	I could use a calculator. Fifteen times 20 equals 300.

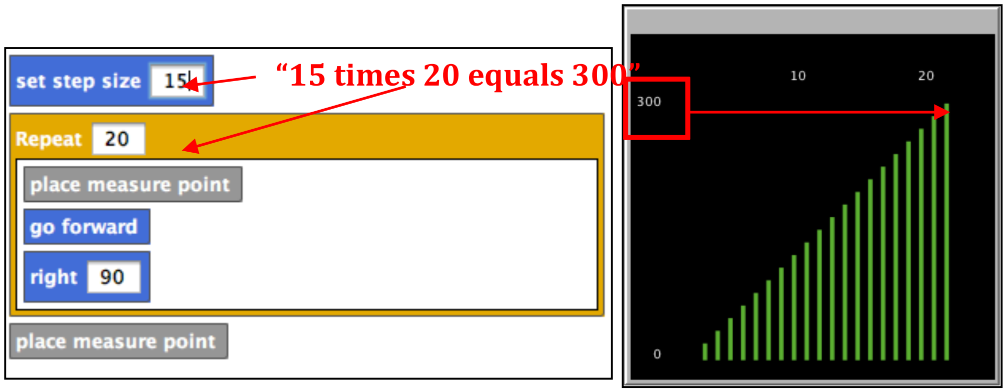


Figure 4: Angelo's ViMAP model

Further into Prediction: Generalizing Motion using a Multiplicative Scheme

Toward the end of Phase II, Emma and researchers wanted to extend the thinking students had done on developing predictive models of motion into more generalizable mathematical forms. The teacher recognized that the formula for Total Distance derived by the class (Number of Steps x Approximate Step Size) was a specialized form of the a multiplicative scheme that also serves as a rate equation: *Distance = Speed x Time*. She told the researchers that she considered this to be a great context for engaging her students in multiplicative reasoning. She explained to the class that this is a "powerful" formula, which can be used to analyzed many real-world situations. She then introduced a "real world" problem, in which students had to of figure out which of two cars, Car 1 or Car 2, traveled further. Car 1 traveled at a speed of 45 mph for 3 hours, while Car 2 traveled at a speed of 35 mph for 4 hours. A sample student's work is shown in Figure 5. As students shared their ViMAP models, we noticed that all of them were able to produce ViMAP models that used appropriate and non-redundant variables. The multiplicative reasoning was evident in students' use of "repeat" and "step-size", as shown in Figure 6, where Car 1 travels 3 (repeat) x 45 (step-size) units, and Car 2 travels 4 (repeat) x 35 (step-size) units

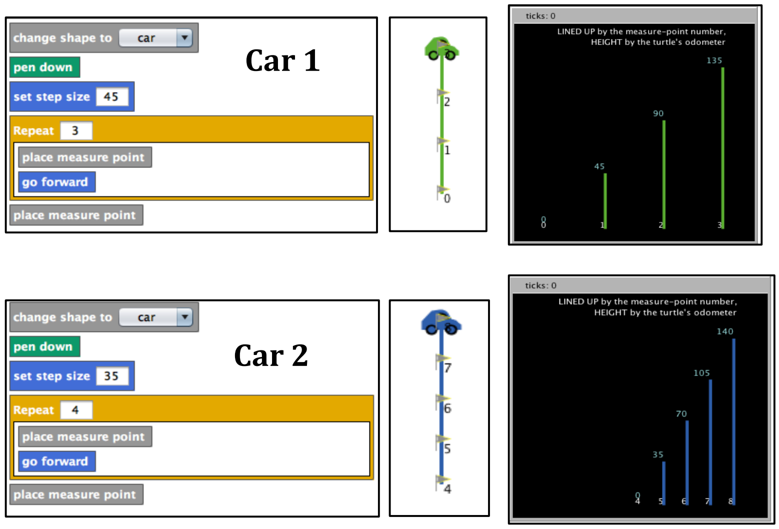


Figure 5: A student's solution to the two-car problem using ViMAP

Co-development of Sociomathematical Norms and Computational Thinking

Our analysis also shows that across the class, there was also an increase in students' ability to compose ViMAP models that accurately represented their data. Students' use of, as well as their skill, at generating accurate ViMAP graphs also increased over Phase II. The growth in students' computational fluency is evident in Figure 6, which shows how students' use of the ViMAP programming commands became increasingly sophisticated as they held their models accountable to the sociomathematical norms throughout the duration of the activities reported in the paper (Phase II). We scored each student's final ViMAP model at the end of each class period in terms of whether they used appropriate variables in their ViMAP code, and whether their graphs represented appropriate element(s) of the phenomenon being simulated using their ViMAP code, each on a scale of 0 – 3. A score of zero meant none of the variables used were appropriate, whereas a score of 3 meant no use of redundant or incorrect variables. The accuracy of the graphs in students' later models were indicative of the appropriate use of the "repeat" command, and order of placement of the “place measure” command. This in turn relied on a conceptual understanding of when to initialize the measurement, and how often the desired measurement had to be repeated in order to generate the graph.

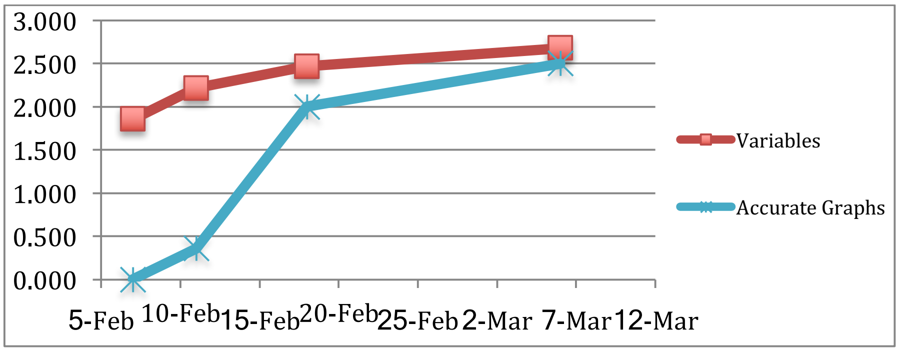


Figure 6: Improvement in Computational Thinking

Why did this improvement happen? We believe that the illustrative cases we presented shows that the development, deployment and refinement of sociomathematical norms led to iterative improvement in the quality of students’ models as progressively more *authentic* representations of the phenomena they were modeling. Emma was an integral participant in this refinement, evident in her push for “accuracy” which was often taken up by students in their modeling work, and became a disposition that was “taken as shared” (Cobb, Wood, Yackel, & McNeal, 1992) by the classroom community. The push to develop “mathematically accurate” measures of distance and “accurate”, in terms of total distance, predictive models in ViMAP resulted in deepening of students’ multiplicative reasoning through the use of ViMAP programming, and this was evident in their use of loops and agent-level variables (No. of Repeat x Step size), as well as a more careful attention to the design of graphs.

Interestingly, students’ facility with ViMAP as a modeling tool progressed in a manner that was initially interpreted by the researchers to be at odds with the epistemic goals of the research. As students first used ViMAP as a way to model their step size values, Emma focused on the computational power of ViMAP programs, constraining ViMAP's function to similar to that of a calculator. Instructing the class as they worked to design programs that model their individual step sizes, she comments that if they put all of their commands in correctly "there is a really neat *trick*, you can look at your graph and it will *add up* your measurements so you can see if you're right". It is clear in this exchange that to Emma, ViMAP is not a modeling tool, as evidenced by her use of the word "trick" to summarize ViMAP's functionality and her focus on ViMAP's ability to quickly compute total distances. This instructional move was initially viewed by the researchers as unproductive, however, it soon became clear that the teacher's explicit focus on graphs – even as a computational tool – was a productive pathway for refining students' computational thinking and modeling practices. In her instruction, Emma framed graphing as a debugging tool, instructing students to "check their graphs" as incorrect graphs were a "clue" that student programs were "missing commands". Programing in ViMAP was no longer seen as extraneous to learning science; rather, the establishment of sociomathematical norms reified the use of ViMAP programming as the *language* of doing science.

Furthermore, we also believe that the teacher's emphasis on using physical and embodied modeling as a way to complement computational modeling and thinking played an important role in the students' take-up of the norms. Cobb and colleagues have argued that sociomathematical norms pertaining to what counts as an acceptable mathematical explanation and justification typically have to be interpretable in terms of actions on mathematical objects that are experientially real to the listening students, rather than in terms of procedural instructions (Cobb, Wood, Yackel, & McNeal, 1992). In our study, by emphasizing embodied modeling as a way to mathematize motion, the teacher facilitated the students' take-up of norms pertaining to "what counts as a good model" of motion, by making ViMAP commands such as "step-size" experientially real to the students. Furthermore, the teacher's focus on the measure command as a way to "see" individual steps enacted by the ViMAP turtle helped students to discretely represent the motion of the turtle agent and correctly interpret the resultant graphs. At the end of phase II, the communicative nature of graphs was privileged, with students remarking that novices unfamiliar with ViMAP as a modeling tool would be unable to interpret turtle enactment as it relates to the phenomena, remarking that if novices "look at the [enactment] it wouldn't be understandable, but if they look at the graph, they will *know*".

Discussion

Sociomathematical Norms Can Integrate Computational Thinking and Science

Our study highlights the *reflexive* relationship between computational thinking, scientific modeling and mathematical thinking when agent-based programming is the computational medium. While this has been noted previously in researcher-led studies (Kafai & Harel, 1991; Papert, 1980; Sengupta et al., 2013), our work here shows that teachers with no background in programming can integrate programming with their existing science curricula by reframing programming as mathematization – in particular, designing measures of change. Furthermore, our study also shows that using agent-based programming as the means to develop these models of change can be supported by the teacher by developing sociomathematical norms around the mathematical quality of these models.

Pragmatically, such forms of integration can be truly synergistic for the K12 science classroom. As we have reported elsewhere (Sengupta et al., 2015), interpreting and constructing mathematical measures (for example, units of measurement and graphs) is a commonly experienced difficulty for students in science classes. Manipulating units is emphasized in statewide standardized assessments, so it is an important learning goal. Agent-based programming can help students overcome these challenges because the activity of programming the behavior of agents requires the learners to define the event in discrete measures. The state of the simulation, at any instant, represents a single event in the form of spatialized representations of agent actions and interactions. To "run" the simulation, these events are repeated a number of times specified by the user. By engaging in iterative cycles of building, sharing, refining, and verifying ViMAP models, students refine their understanding of what actions and interactions of agents represent an "event," which are then displayed on graphs. This provides students with the opportunity to explore different kinds of units, and see their simulation measured in those units. Teacher initiated *sociomathematical* norms, such as the ones reported in this study, when taken up in student work through joint action, can help students harness and deploy the epistemic and representational power of agent-based computing as the "language" for doing science. New literacies such as computational modeling and programming can thus be meaningfully and seamlessly negotiated with day-to-day needs in the science classroom.

Methodological Concerns: Teacher Voice and Conceptual Dissonance in Researcher-Teacher Partnerships

Design-based researchers have recently begun advocating for greater teacher voice and agency in research studies, which in turn reframes studies as researcher-teacher partnerships (Severance, Penuel, Sumner & Leary, 2016). Our study is certainly an example where teacher voice often led the direction of research; but it also raises an important methodological and epistemological question: how should we address conceptual dissonances between the researchers and the teachers? For example, in our study, the teacher's framing of "accuracy" - i.e., students' models must be "mathematically accurate" - was largely based on her intuitive conceptualization of the term. Let us now imagine answering this question as educational researchers and epistemologists. “Accuracy” will take on a very different meaning, and perhaps even have a negative connotation - because an essential characteristic of models, according to the epistemologists of science, is that they are incomplete. In fact, a few months later, the teacher did introduce the notion of incompleteness (albeit in her own language, and in a different context) – in Phase III, while modeling ecological interdependence. The notion of accuracy, though, lingers throughout the academic year.

We will take up this issue in more detail in a different paper. But we do want to raise the following question here: what should we do in such situations? Should we have intervened and coached the teachers about the professional vision of scientists and epistemologists about accuracy and incompleteness of models? This study is an example where we did not intervene to bridge conceptual dissonance on this issue. Our decision stems from the fact that researchers must fundamentally position teachers as the director of the partnership – rather than at an equal footing with the researcher. An equitable partnership may not be one in which everyone has equal say. Instead, an equitable partnership in educational computing research must seek to support teachers in voicing (and re-voicing) computation from their own perspectives, with curricular mandates and classroom constraints in mind.

As Heidegger famously remarked, the essence of technology is nothing technological (Heidegger, 1954). Rather, it is the “frame” in which technology lives – its lifeworld of human experience – that defines it. Unfortunately, researchers in educational computing – in particular, programing languages for children – have traditionally not engaged with the issue of curricular integration from the perspective of K-12 teachers. Research studies in this field (including some of our earlier work), therefore, largely carry out a strong interventionist agenda where teacher voice is often overshadowed by the researchers. In contrast, we have come to see the K-12 public school classroom as a complex, interdependent system, where teachers, students, curricula and curricular mandates – must all be considered alongside one another, especially if we set out to integrate any new literacy and/or technology with the classroom. So, if our goal is to make programming and computational modeling ubiquitous in the K-12 science classroom, we posit that researchers and designers of programming languages for the K-12 classrooms must learn to see the world through the eyes of the teachers, especially when it involves conceptual dissonance between researchers and teachers. It is through carefully studying the unfolding of such dissonances over longer periods of time (i.e., not a short intervention study), especially when teachers are working with new technologies and literacies (such as programming and computational modeling), that we (as researchers) will learn to design technological and activity systems that will be aligned with the perspectives of the teachers, and therefore, have a greater chance of becoming a mainstay in their classrooms.

References

Cobb, P., Wood, T., Yackel, E., & McNeal, B. (1992). Characteristics of classroom mathematics traditions: An interactional analysis. *American Educational Research Journal*, *29*(3), 573-604.

diSessa, A., Hammer, D., Sherin, B., & Kolpakowski, T. (1991). Inventing graphing: Children's meta-representational expertise. *Journal of Mathematical Behavior*, *10*(2), 117-160.

Dickes, A., Sengupta, P., Farris, A.V., & Basu, S. (2016). Development of Mechanistic Reasoning and Multi-Level Explanations in 3rd Grade Biology Using Multi-Agent Based Models. *Science Education, 100*(4), 734–776.

Dykstra Jr, D. I., & Sweet, D. R. (2009). Conceptual development about motion and force in elementary and middle school students. *American Journal of Physics*, *77*(5), 468-476.

Farris, A.V., & Sengupta, P. (2014). Perspectival Computational Thinking for Learning Physics: A Case Study of Collaborative Agent-based Modeling. In: *Proceedings of the 12th International Conference of the Learning Sciences.* (ICLS 2014), pp 1102 - 1107.

Ford, M. J., & Forman, E. A. (2006). Redefining disciplinary learning in classroom contexts. *Review of research in education*, *30*, 1-32.

Giere, R. N. (1988). *Explaining science: A cognitive approach.* Chicago: University of Chicago Press.

Heidegger, M. (1954). The question concerning technology. *Technology and values: Essential readings*, 99-113.

Kafai, Y., & Harel, I. (1991). Learning through design and teaching: Exploring social and collaborative aspects of constructionism.

Lehrer, R. (2009). Designing to develop disciplinary dispositions: modeling natural systems. *American Psychologist*, *64*(8), 759.

Lehrer, R., Schauble, L., & Lucas, D. (2008). Supporting development of the epistemology of inquiry. *Cognitive Development*, *23*(4), 512-529.

Lehrer, R., & Schauble, L. (2000). Developing model-based reasoning in mathematics and science. *Journal of Applied Developmental Psychology*, *21*(1), 39-48.

Lehrer, R., Schauble, L., Strom, D., & Pligge, M. (2001). Similarity of form and substance: Modeling material kind. Cognition and instruction: Twenty-five years of progress, 39-74)

Lehrer, R., & Schauble, L. (2006). *Cultivating Model-Based Reasoning in Science Education*. Cambridge University Press.

McClain, K., & Cobb, P. (2001). An analysis of development of sociomathematical norms in one first-grade classroom. *Journal for Research in Mathematics Education*, *32*(3), 236-266.

Next Generation Science Standards (2013). www.nextgenscience.org/next-generation-science-standards.

Papert, S. (1980). *Mindstorms: Children, computers, and powerful ideas*. Basic Books, Inc.

Pickering, A. (1995). *The mangle of practice: Time, agency, and science.* Chicago: University of Chicago Press.

Sengupta, P., Dickes, A., Farris, A. V., Karan, A., Martin, D., & Wright, M. (2015). Programming in K-12 science classrooms. *Communications of the ACM*, *58*(11), 33-35.

Sengupta, P., & Farris, A. V. (2012). Learning kinematics in elementary grades using agent-based computational modeling: a visual programming-based approach. In: *Proceedings of the 11th International Conference on Interaction Design and Children* (pp. 78-87). ACM.

Sengupta, P., Kinnebrew, J. S., Basu, S., Biswas, G., & Clark, D. (2013). Integrating computational thinking with K-12 science education using agent-based computation: A theoretical framework. *Education and Information Technologies*, *18*(2), 351-380.

Sherin, B., diSessa, A. A., & Hammer, D. (1993). Dynaturtle revisited: Learning physics through collaborative design of a computer model. *Interactive Learning Environments*, *3*(2), 91-118.

Severance, S., Penuel, W. R., Sumner, T., & Leary, H. (2016). Organizing for Teacher Agency in Curricular Co-Design. *Journal of the Learning Sciences*, (just-accepted).

Vee, A. (2013). Understanding computer programming as a literacy. *Literacy in Composition Studies*, *1*(2), 42-64.

Wilensky, U. (1999). {NetLogo}.

Wilkerson-Jerde, M., Wagh, A., & Wilensky, U. (2015). Balancing curricular and pedagogical needs in computational construction kits: Lessons from the deltatick project. *Science Education*, *99*(3), 465-499.

Wing, J. M. (2006). Computational thinking. *Communications of the ACM*, *49*(3), 33-35.

Yackel, E. and Cobb, P. (1996) Sociomathematical norms, argumentation, and autonomy in mathematics. *Journal for Research in Mathematics Education*, *22*(4), 390-408.