

# Peeking into the Black Box: Modeling as an Epistemic Tool for Building Student Disciplinary Knowledge and Scientific Inquiry Skills

Kam D. Dahlquist\*, Department of Biology, Loyola Marymount University, Los Angeles, CA 90045  
Melissa L. Aikens, Department of Biological Sciences, University of New Hampshire, Durham, NH  
03824  
Joseph T. Dauer, School of Natural Resources, University of Nebraska-Lincoln, Lincoln, NE, 68503  
Samuel S. Donovan, Department of Biological Sciences, University of Pittsburgh, Pittsburgh, PA 15260  
Carrie Diaz Eaton, Digital and Computational Science, Bates College, Lewiston, ME 04240  
Hannah Callender Highlander, Department of Mathematics, University of Portland, Portland, OR 97203  
Kristin P. Jenkins, BioQUEST, Boyds, MD 20841  
John R. Jungck, Department of Biological Sciences, University of Delaware, Newark, DE 19716  
M. Drew LaMar, Department of Biology, College of William and Mary, Williamsburg, VA, 23187  
Glenn Ledder, Department of Mathematics, University of Nebraska-Lincoln, Lincoln, NE 68588  
Robert L. Mayes, College of Education, Georgia Southern University, Statesboro, GA 30460  
Richard C. Schugart, Department of Mathematics, Western Kentucky University, Bowling Green, KY  
42101

\*Corresponding author:

Kam D. Dahlquist  
Department of Biology  
Loyola Marymount University  
1 LMU Drive, MS 8888  
Los Angeles, CA 90045 USA  
Tel: 310-338-7697  
Fax: 310-338-5317  
[kdahlquist@lmu.edu](mailto:kdahlquist@lmu.edu)

Article type: Essay

Character count (with spaces): 54,404 (without title page or references); 77,203 (entire document)

Running title: Peeking into the Black Box

Keywords: models and modeling, process of science, quantitative skills, introductory biology,  
undergraduate

## Abstract

Models and the process of modeling are fundamental to the discipline of biology, and therefore should be an important component of undergraduate biology courses. Engaging students with models and modeling is an epistemic approach which can provide opportunities to use widely shared scientific practices to make and evaluate knowledge claims. Modeling can support effective pedagogy because it provides students opportunities to ask questions, seek evidence, perform analyses, and construct arguments that can increase students' understanding of the scientific process and biological content. Modeling also affords opportunities for anchoring course content in real-world problems and facilitating metacognition. Models can be found in any introductory biology curriculum, and we present suggestions to help instructors leverage models and modeling for greater learning in the classroom. We use the Hardy-Weinberg Equilibrium model to illustrate how students can progress from using models as a "black box", to looking inside to explore the why and how of the observed behaviors, and finally to propose models of their own. We propose that even small shifts in the way models and modeling are used in the classroom could provide increased understanding of key biological concepts, insights into realistic scientific inquiry, and opportunities to build quantitative and communication skills.

## Overview

The central role of models and modeling in scientific practice should be reflected in the way we teach science. Scientists use models to study complex systems, make predictions, test ideas that are experimentally difficult or impossible to test, develop conceptual frameworks, and generate causal relationships (Box 1, Table 1; Odenbaugh, 2005; Svoboda and Passmore, 2011). Biologists' use of models to communicate ideas and explore theories is an integral part of the scientific process (Tomasi, 1988; Gilbert, 1991; Lander, 2010; Jungck, 2011). Although, there are many types of models, including physical, computational, diagrammatic, and quantitative (Bryce et al., 2016; Frigg and Hartman, 2017), all models share the goal of representing the components of a system and the relationships between those components. Using a variety of model types to represent different aspects of the same system can provide insights for researchers and students alike. For example, physical models built from ball and stick modeling kits are useful for understanding molecular structures, such as the primary structure of a peptide or the secondary structure of DNA (Figure 1A; Cooper & Oliver-Hoyo, 2017). Pairing these physical models with computational models that can be used to visualize properties such as hydrophobicity and charge (e.g. Jaswal et al., 2013; FirstGlance in Jmol, among others) or sequence conservation (e.g. ConSurf; Ashkenazy et al., 2016) can increase understanding of the complex intra- and intermolecular interactions of macromolecules, such as regulatory transcription factors binding to DNA. As computational power and access to data increases, biologists are able to employ quantitative models more frequently and to greater effect. For example, the Critical Assessment of protein Structure Prediction (CASP) has seen substantial improvement in models used to predict protein structures (Moult et al., 2018).

For students, working with models can lead to a deeper understanding of key biological concepts, provide practice in realistic modes of scientific inquiry, and build both quantitative and communication skills (Windschitl et al., 2008; Lehrer and Schauble, 2005; Garfunkel and Montgomery, 2016). Indeed, national reports and proposed standards for improving biology education have advocated for the

development of competencies and skills related to using models and modeling (AAMC-HHMI, 2009; AAAS, 2011; NGSS, 2013; College Board, 2015). Engaging with multiple representations of a system provides students with repeated exposure to the biological concepts underlying the components and relationships in a system (Bryce et al., 2016). The process of building models of a complex system is valuable because it helps a student evaluate which are the key characteristics and relevant interactions in a system (Wimsatt, 1987; Nesserian, 1999; Sins et al., 2005). In support of quantitative modeling, it has been shown that teaching students quantitative modeling both increases and refines their conceptual understanding of biological processes (Svoboda & Passmore, 2011; Schuchardt and Schunn, 2016).

### *Who we are and our goals for this essay*

We, the authors, are an interdisciplinary group of biologists, mathematicians, mathematical biologists, and education researchers who came together for a working group at the National Institute for Mathematical and Biological Synthesis (NIMBioS), organized by the leadership team from the Quantitative Undergraduate Biology Education and Synthesis project (QUBES, 2017b; Donovan et al., 2015), to address the challenges of teaching modeling in general and introducing students to quantitative modeling in particular. The working group has provided us the opportunity to explore the education research on modeling, share how modeling is applied and taught in our various disciplines, and examine our individual teaching experiences for best practices. There is an extensive education literature on models, modeling, and model-based reasoning. Our discussions were informed by close readings of selections from this literature, cited in this essay, but these are not meant to be taken as an exhaustive review. Instead we wish to share insights distilled from our discussions about how we might improve our approach to teaching with models in biology courses, especially introductory biology, without substantial changes to the curriculum. In this essay we have three goals:

1. To explore the epistemic value of using models and modeling to increase students' understanding of biological concepts;
2. To link established pedagogical approaches to the teaching of models and modeling; and

3. To suggest ways that scaffolding modeling in introductory courses provides a foundation for students to perform quantitative modeling in both introductory and advanced-level courses.

We hope these goals inspire you to enlist your students in “peeking into the black box” by emphasizing models, modeling, and model-based reasoning more frequently in your biology courses.

## Epistemic Value of Models and Modeling

### *Describing models*

Our working definition is that a *model* is a simplified representation of real-world objects and their mechanistic or functional relationships, constructed for a purpose, such as understanding or making predictions about a real-world phenomenon (Box 1; Table 1; Frigg, 2002; Lehrer and Schauble, 2010; Eaton et al., 2018). The statistician George Box is credited with the observation, “All models are wrong, but some are useful.” (Box, 1976). Clearly, models are incredibly useful in science, but Box’s quote serves as a reminder that no model represents all aspects of a system’s components and relationships. A model is an inherently incomplete representation of a system because in order to be simple enough to be useful, a model is built based on assumptions that foreground certain aspects of a system while leaving others out entirely. For example, in a simple SIR (Susceptible, Infectious, Recovered) epidemiology model (Kermack and McKendrick, 1927), it is assumed that there is a constant recovery rate for infected individuals. This is not how diseases work in the real world, but despite this obvious flaw, SIR models provide surprisingly useful information for predicting the pattern of disease outbreaks that can be used to inform the timing and distribution of health care responses (Keeling and Danon, 2009). However, it is important to be aware of the assumptions and omissions of a model and how they influence the output. A more complex model can be built that reduces the number of assumptions and incorporates more variables, but doing so constrains the model and limits its flexibility. Thus, both the value and limitation of models is that they simplify systems. When a student can explain how the assumptions of a model can

affect model outcomes, she is moving beyond treating the model as “content” to be memorized and is progressing toward model-based reasoning.

One way to compensate for simplifying assumptions is to use multiple types of model representations to fully explore a system. Model representations can be experiential (physical manipulatives, animations/simulations, experiments; Figure 1A), visual (schematics, diagrams, flowcharts; Figure 1B), verbal (hypotheses, predictions, descriptions, assumptions; Figure 1C), numerical (data tables, graphs; Figure 1D), or symbolic (equations, formulas; Figure 1E) (Simundza, 2006; Eaton et al., 2018). Using multiple representations of a system is particularly relevant as a pedagogical approach to support student understanding. Each model provides insight into some aspects of the system and compensates for the shortcomings of other models (Svoboda and Passmore, 2011; Bryce et al., 2016). In addition, moving between different model representations provides an opportunity for conceptual change (Nessarian, 1999). For example, genes can be represented in different ways in different models. A molecular biologist might use a visual gene diagram in a model of the mechanism of gene expression, whereas a population geneticist might incorporate genes symbolically into a mathematical model for understanding how evolutionary processes act on populations (Figure 1A, 1E; Marbach-Ad & Stavy, 2000; Dauer et al., 2013). Both of these representations of genes are abstractions, including some details and leaving others out appropriate to the desired application and the relevant biological theories. In these two examples, the modelers have organized their knowledge of the system, focusing their attention on the aspects of most relevance to the question they are asking. Neither model encompasses all aspects of the phenomenon of “genes” because not all models are appropriate for all purposes (Table 1; Odenbaugh, 2005; Frigg and Hartman, 2017). When a student connects these two models and can describe the utility of each, he has enhanced his disciplinary knowledge.

### *Engaging in the modeling process*

Using multiple model representations is part of the process of modeling. Our working definition is that *modeling* is an iterative process in which a model is proposed, explored, validated, and refined

(Box 1; Figure 2; Louca and Zacharia, 2012; Garfunkel and Montgomery, 2016). Engaging in modeling both nurtures and reflects a growing understanding of specific phenomena and overarching concepts in students (Nesserian, 1999; 2002). Proposing models to organize information, focusing on potentially relevant features, and communicating with others about ideas allows the student to explore the “goodness of fit” between her mental model and scientific models, which, in turn, reflects her understanding of a biological concept and the real world (Ifenthaler, 2011). Teaching the process of modeling is the difference between having students sketch a picture of a gene (which often results in drawings of an X-shaped chromosome or a double helix) versus having them propose a gene model that explains the interactions, relationships, and observed properties of gene expression. The resulting diagram of a gene model that has a promoter, transcription start site, open reading frame, and terminator begins to explain the mechanism of gene expression (Figure 1B). This pedagogical difference can be described as “models *of* versus models *for*” (Gouvea and Passmore, 2017). When students are engaged in “modeling *for*” understanding, they learn biology more deeply.

Like any skill, learning how to effectively propose models requires instruction and repetition that comes by generating different models with application to different biological problems. Students engaging in modeling for the first time will need support moving from the concrete real world to an abstract representation (Hestenes, 1992; Lehrer and Schauble, 2010; Mulder et al., 2016). In order for students to effectively propose models, they must be able to envision features of the real world in an abstract form that can be measured (Kline, 1980). Initially students may need their abstract representations to closely resemble the real world. With practice, students can move to more abstract representations that focus on just the relevant components of the phenomenon under study (see Lehrer & Schauble, 2005 for additional references), including quantitative models. The development of “representational competence”, which is the ability to interpret and move between the abstract representations of a phenomenon (diSessa, 2004; Lehrer and Schauble, 2010; Sim and Daniel, 2014), is a reflection of an increase in model-based reasoning and deeper disciplinary understanding.

While it may sound daunting to provide novice modelers with the support they need to develop these complex cognitive and scientific skills, we suggest that effective teaching of these skills requires a change in approach, rather than a substantial change in curriculum. The process of modeling can begin as simply as sketching an explanation for an observed phenomenon for the purposes of discussion (Dauer et al., 2013; Long et al., 2014; Box 2). Such a model meets multiple educational needs simultaneously. Students organize the information they have about the phenomenon mentally, and in sketching and discussing the model, begin to refine it and identify gaps in their knowledge. Meanwhile, faculty receive a glimpse into students' current understanding of the phenomenon (Pearsall et al., 1997; Speth et al., 2014). For example, sketching a visual model of how odor molecules move through space gives students an opportunity to organize their understanding of the phenomenon, communicate their ideas to others for discussion, as well as suggest how to test that idea (Nesserian, 1999; Gilbert, 2004). This initial model is an opportunity to identify key components of the system, and decide if it is critical to one's purpose to quantify the relationships.

It is important to recognize the enormous cognitive shift required of students as they move from real world biological systems to abstract model representations, and to support or scaffold that transition (Lehrer and Schauble, 2010). To support this transition, it can be helpful to represent a model in multiple modalities (experiential, visual, verbal, numerical, and symbolic, Figure 1, cf. Eaton et al., 2018). Having students use multiple representations of a model which complement one another can support disciplinary learning (Ainsworth, 1999), as different representations highlight different aspects of a phenomenon. Importantly, the initial model should be considered a starting point, and should be repeatedly revisited. Revising and refining the model based on new information and a deeper understanding of the phenomenon is an important part of modeling (Latour, 1999; Hogan and Thomas, 2001).

Another key aspect of modeling is returning to the biological system and asking if the outputs from the model make biological sense. The output of a mathematical model in particular must be returned to the original biological question to consider whether the modeling outputs make sense in the



biological context (Schuchardt and Schunn, 2016), empirically testing the model against the observations of the phenomena to determine if the model fits the observed data (Windschitl et al., 2008; Schwarz et al., 2009). This is a key moment where conceptual change can occur. As students try to reconcile their explanation of the phenomenon, their model, with the actual event, they must articulate how the outcomes of the model support understanding of the phenomenon (Mayes et al., 2013a). Explaining the connection is an opportunity for intellectual growth, and additional growth is afforded when this step is followed by an opportunity to revise and refine the model. Shortcomings of a model can be addressed, or different aspects of a model might be foregrounded. Students can also refine their models by comparing them with models from their peers and with established models of the phenomenon. This provides the opportunity for students to develop the important soft skills of communication and collaboration, while increasing their deep understanding of biological processes.

### *Quantitative modeling and the process of science*

The relationship between various scientific models and quantitative models bears some exploration at this point. Many models can be represented symbolically as mathematical equations, but a mathematical equation needs to be grounded in a particular context with meaning to be considered a mathematical model (Frigg et al., 2012). Further, one of the strengths of mathematical modeling is that, in many cases, the same mathematical equation is appropriate to describe interactions or relationships of components of different models. When one equation can describe multiple phenomena, the modeler can use information from other phenomena that can be described by the same mathematical equation to generalize about the behavior of the new system to which the equation is applied (Nesserian, 1999). For students, recognizing the similarities between models described by mathematical equations reduces the cognitive load of information-rich biology courses. For example, an exponential function can describe both population growth and radioactive decay. In both cases, the use of the exponential function provides critical information about the assumptions and behavior of the system. In contrast, statistical models are

mathematical models derived from data that allow users to evaluate the probability of an outcome when the entire system cannot be studied, and instead, samples are examined.

Computational models include simulations and other software-based modeling approaches, such as agent-based models. Computational models are frequently based in complex mathematical or statistical descriptions of a phenomenon, yet have a unique epistemic role to play in modeling, especially for novices. Modifying mathematical or statistical models to explore a biological phenomenon requires students to grapple with both the biological and quantitative aspects simultaneously, creating additional challenges for learners. However, computational models allow students to explore biological phenomena without being intimidated by the underlying mathematics or statistics. Students can use these “black box” constructions to explore extremes, null models, and the effects of stochasticity by comparing differing outcomes generated by repeatedly running the model (Wilensky and Reisman, 2006). For the purposes of this discussion, we will no longer focus on the distinction between mathematical, statistical, and computational models and will refer to all as either “mathematical” or “quantitative” models interchangeably.

Figure 2 is meant to help show students the epistemic value of mathematical modeling. We have used model-based reasoning (Windschitl et al., 2008) as a framework to show the parallels between an experimental approach and a mathematical modeling approach to the process of science. In this framework, to address a biological problem, we must identify the problem in its biological context, organize the relevant information, ask questions, and formulate hypotheses (Figure 2, top). We then seek evidence, in this example through experiments or quantitative modeling (Figure 2, middle, discussed below). Finally we analyze the results, constructing an argument for a biological interpretation of the results within a larger disciplinary conceptual framework that is then subjected to review by the scientific community (Figure 2, bottom). Feedback from other scientists leads to a reframing of the biological problem itself, kicking off the next iteration of the entire process.

Windschitl and colleagues (2008) stress the idea of viewing science as the process of creating and refining explanatory models of the world. Model-based reasoning includes the practice of seeking multiple lines of evidence (e.g., mathematical and/or experimental) in order to choose one hypothesis from a set of competing hypotheses. Thus, as can be seen in Figure 2, mathematical models and experimental models are different ways to construct evidence to test a hypothesis, with parallels even within the individual steps of each (Servedio et al., 2014). Within this framework, experimental research (Figure 2, right) and mathematical modeling (Figure 2, left) are two parallel tracks that can be pursued when seeking to understand a phenomenon. While we highlight experiments in Figure 2 (right), we note that observational field studies, evolutionary reconstructions, and meta-analyses are other modes of investigation carried out by biologists. There is a philosophical argument to be made that experimental research could be considered modeling, since the experimenter and modeler both go through a process of defining a system for study and identifying variables and parameters relevant to the study. However, the main point we wish to make here is to draw attention to the use of quantitative modeling as an opportunity for students to engage in the process of science. Even if it is not feasible to offer a hands-on experimental science experience in class, it is entirely possible to give students experience with the scientific process using quantitative models. Modeling is a critical tool in research and including it in introductory course instruction emphasizes this.

Engaging in quantitative modeling is challenging for novices since it requires moving from the biological realm to the symbolic mathematical realm, quantifying components and relationships in a more abstract way. Moving to an abstract representation of a real phenomenon is made even more challenging by the language barrier between mathematical and biological disciplines. Mathematicians and biologists may use entirely different terms to describe the same function (e.g. the first order autonomous differential equation versus specific application areas, such as Newton's Law of Cooling, Von Bertalanffy growth, or the first-order kinetic model of a reaction product), making it more difficult for novices to bridge the gap between biological and mathematical representations (Feser et al. 2013). Often in biology, students are

not comfortable with mathematical descriptions of systems. Proposing a model in the form of a verbal description or visual schematic can provide a transition between the real world and the symbolic mathematical representation. This first allows students to explore the biological system and identify the key components and their relationships in more familiar descriptive terms before translating it into mathematical language. In the SIR model, students can easily relate to the three disease states (Susceptible, Infectious, Recovered) and represent them as variables for quantitative analysis (Jungck, et al. 2010). Performing the modeling process (as opposed to learning a model as “content”) also answers student questions about “why do I need to know this?” (Schwarz and White, 2005).

## Linking Established Pedagogical Approaches to Modeling

Members of our working group have used models in the classroom to accomplish a variety of learning outcomes for our students, ranging from a clearer understanding of a biological process to the development of mathematical modeling skills (Boxes 2 & 3). The discussion of our collective experiences has given us a more practical understanding of the problems that students and faculty face when including modeling as part of the curriculum. Because modeling, and specifically quantitative modeling, can be a complex and lengthy process which requires time not often available in the biology curriculum, we do not expect students to engage in the entire modeling process in every course (Figure 2 left, top to bottom). Instead, we engage students frequently in more granular “modeling activities” (individual arrows of Figure 2) that can be supported in every class as part of a longer learning trajectory across their undergraduate experience (Eaton et al., 2018).

There are many ways to leverage the existing presence of models and modeling in the curriculum to achieve greater learning outcomes. Integrating modeling specific methods with evidence-based pedagogical practices, such as engaging students in compelling, real world problems (Box 3; Schwarz et al., 2009) can be extremely productive. Using relevant, real biological problems provides students with a “need to know” that drives their interest in the problem (e.g., Dohn et al., 2009), motivating students as they struggle to learn new modeling skills and new disciplinary information (Hidi and Harackiewicz,

2000). In addition, real biological problems provide an opportunity for students to engage in both the scientific process and the modeling process by asking questions about a biological phenomenon (Figure 2, top). The ability to ask good scientific questions is an important part of the scientific process, and is another opportunity for students to increase their understanding of this process as well as biological concepts (Jungck, 1985; Peterson and Jungck, 1988; Windschitl et al., 2008; Rothstein and Santana, 2011). This fundamental process-of-science skill requires practice and is worth the time to develop. An iterative process, in which students pose questions, sketch models that align with those questions, and then move between the question and sketched model, refining both, will help students (and their instructors) identify where they are missing information (Pearsall et al., 1997; Dauer et al., 2013; Long et al., 2014; Speth et al., 2014). When students produce a model, they can then discuss what quantitative aspects are included, or should be included, in the model and for what purpose (Weisstein, 2011).

Shifting between a qualitative verbal or visual model and a quantitative mathematical model (mathematization, Figure 2, left track) can be particularly challenging for students. The first steps in mathematizing a problem involve identifying the variables, a unit measure for each variable, and attributes of each variable that help determine covariation between variables (Thompson, 2011). One way to help students with moving from a qualitative to a quantitative model is to use modeling language about variables, assumptions and limitations (Brewer, 2008) and discuss concepts such as model use and utility (Windschitl et al., 2008). In addition to moving between multiple representations, it is vital to connect the biological phenomenon with the qualitative and quantitative interpretation of these models. The biological context should serve as the unifying link between models, evaluating trends, and making predictions based on models (Mayes et al., 2013b). By changing how we think about presenting this information as part of a lecture or activity, we can achieve the goals of raising student awareness and understanding of models, increasing the recognition of the epistemic value of modeling for both learning and research, and giving them the language and skills to develop proficiency in modeling.

These practices, in turn, can promote student metacognition about both modeling and disciplinary learning. The goal of having students reflect about using models is to help them develop an awareness and evaluation of their thinking (a.k.a., metacognition) about models when they encounter new scientific problems. Some have defined this level of awareness about modeling as “metamodeling”, which is the ability to be metacognitive about the process of modeling (Boxes 2 & 3; Schwarz and White, 2005; Papaevripidou et al., 2007). Metamodeling can improve students’ understanding of practices like predicting, observing, and explaining phenomena (Barab et al., 2000; Schwarz and White 2005; Sins et al., 2005) and the ability to make mechanistic explanations (Fretz et al., 2002; Louca and Zacharia, 2012). Importantly, metamodeling enhances students’ abilities to regulate their own learning (Papaevripidou and Zacharia, 2015). As students gain awareness of where they are relative to a learning progression of modeling (Schwarz et al., 2009), they can be more aware of how they are using models to address biological problems. In the section that follows, we flesh out these recommendations in the context of the commonly taught Hardy-Weinberg Equilibrium (HWE) population genetics model (Soderberg and Price, 2003).

## Peeking into the Black Box: the Hardy-Weinberg Equilibrium Model

The Hardy Weinberg Equilibrium (HWE) is usually introduced in the “evolution section” of introductory survey courses along with the forces that drive evolution: natural selection, nonrandom mating, genetic drift, gene flow, and mutation. Although we refer to HWE as an equation, it is a null model representing how a simple biological system functions if no evolutionary forces were to act on a population. Already, this presents three cognitive challenges for students. First, students are typically not familiar with the use and value of a null model to explore a biological phenomenon. A common response to HWE from thoughtful, puzzled, and somewhat frustrated students is that HWE never represents reality, so what good is it? Second, at the introductory level, students are often struggling with genetic concepts like “gene” and “allele”. Exploring organisms at the abstract genetic level is a new way of thinking and students’ application of these ideas is often muddled (Speth et al., 2014). Finally, HWE plunges students

into population thinking, which is yet a different level of biological abstraction than thinking about individual organisms. Many misconceptions about evolution are revealed when students extrapolate from a specific example of evolutionary forces at work at the individual level to the resulting evolutionarily meaningful population-level effects. Add to this a modicum of math anxiety, and it becomes clear why HWE, despite its simplicity, presents such a challenge to students.

This set of challenges also provides numerous opportunities to engage students actively in a deeper learning of evolution and population thinking by using models as pedagogical tools. HWE is one of the few mathematical equations to which almost all students are exposed in introductory biology (Box 3). If framed as a simple model of a complex system (Table 1), and used in conjunction with computational models and experiential manipulative models, it can become more accessible, which may help students master core biological concepts. First, to address the challenge of students not understanding the value of a simple, null model in exploring dynamic systems, a pedagogical approach could be to draw parallels between mathematical modeling and scientific experiments (Figure 2). Just as scientists remove extraneous confounding variables in experiments to isolate causative forces and effects, modelers make simplifying assumptions to remove extraneous variables and factors so they can attribute the resulting model behavior to the variables and relationships of interest. Null models can act as controls for the behaviors of complex systems. The ability to predict what would be observed if nothing happened allows researchers to detect when something actually has happened, similar to a negative control in experimental systems. The observed variation from the null model predictions can indicate the forces at work in the system, the targets of those forces, and suggestions for how to study the outcomes. Students readily grasp that the HWE does not accurately represent real world systems, which leads to opportunities to discuss how HWE fails to represent the system, and how the model would need to be altered to more accurately reflect the impact of evolutionary forces. The ability to represent the system in a simple, abstract mathematical representation becomes an efficient and convenient way to discuss complex system behaviors without getting bogged down in the details.

Next, students can explore the concepts of genetic change and evolution with various experiential manipulative models which focus on the behavior and outcomes of an individual in a system. There are many physical models and paper and pencil activities that demonstrate the fluctuations of alleles using small populations (Jungck et al., 2010; Froehlich and London, 1996; Barton, 2000). These activities usually have students simulate the passage of several generations by hand, and generally result in a random distribution of alleles and genotypes. If the results of the manipulative models are compared to the outcomes predicted by the HWE null model, students can see the variation between the null model and their results. When there is variation, an assumption of the HWE model has been violated. In activities where there is an absence of all other evolutionary forces, this variation must be the result of a small population size. This is an opportunity for students to stretch their conceptual understanding from the individual level (or very small population) toward population thinking. The next step in this process is to provide an opportunity to explore allele frequencies in large population, typically through computational models.

Computational models overcome the time constraints of manipulative models and allow students to explore changes in larger populations over hundreds or thousands of generations. They rapidly generate multiple outcomes under conditions that satisfy the assumptions of the HWE null model and in the presence of various evolutionary forces for comparison. A variety of these types of models are freely available (Table 2; Figure 1A). Observation of the repeated emergence of patterns can help students understand how random events at the individual level generate predictable evolutionary outcomes at the population level, supporting the cognitive shift toward population thinking. These patterns can also address genetic misconceptions such as the expectation that dominant alleles will tend to be more frequent in a population. Indeed, this very misconception is what prompted Hardy to generate his version of the Hardy-Weinberg Equilibrium model (Hardy, 1908). Shifting between computational models and the HWE equation provides students with multiple representations of the concepts. When using



computational models, often the algorithms generating the patterns are “black boxes” to students.

Relating these models to the HWE equation provides an entry point to complex calculations.

Students can also validate the HWE model by making predictions and explaining outcomes by testing the model with data collected from existing populations. Starting with a pool of F0 genotypes that are in Hardy-Weinberg equilibrium, students can explore F1 pools and ask if evolutionary forces impacted the allele frequencies observed in each pool. For example, a much smaller number of aa genotypes and much larger number of AA genotypes is consistent with a population in which the homozygous recessive individuals fail to survive to adulthood. For some courses it may be appropriate to introduce students to the Chi-square test to quantify the probability that a given F1 population is in HWE. In doing so, we are layering the Chi-square statistical model on top of the HWE null mathematical model. In terms of model validation, it is important to note that when one compares the observed F1 genotypes to the expected HWE values, one is checking to see if the experimental population is in HWE, but when one compares the expected HWE values to the observed population data, one is also validating the assumptions of the mathematical model.

In the example above, the recognition that the genotype frequencies in the F1 generation do not fit the HWE null model leads to the biological explanation that evolutionary forces may be at play. However, it should also be recognized that this is an evaluation of the assumptions of the mathematical model and that an appropriate revision of the model could improve its explanatory power. This calls for students to cross the threshold into creating a new model through refinement of an existing model to meet new criteria or to apply it to a new situation. With a better understanding of the mathematical modeling process, students can thoughtfully explore the consequences of adding a third allele or a second locus, or explore the outcomes of various adaptive landscapes. These consequences could be reasoned through and tested with the PopGen simulator or Biological ESTEEM Project modules (Table 2). We recognize that this may be beyond the scope of a typical introductory biology course, but include this example here to show where this learning trajectory is leading. The modeling activities described in this essay are a part of

a learning progression from model description and exploration, to model development and refinement. In this process, students shift from seeing mathematical models as a “black box” to a “glass box”, exploring the why and how of the observed behaviors. Finally, by developing and refining their own mathematical models, they are operating with “no box”, having learned skills that can be transferred to new biological problems ([http://bioquest.org/esteem/Intro\\_to\\_ESTEEM.pdf](http://bioquest.org/esteem/Intro_to_ESTEEM.pdf)).

## Repeated Practice: Scaffolding Quantitative Modeling Throughout the Curriculum

While we have explored the HWE model in this essay to demonstrate how to use models as pedagogical tools, we do not wish to imply that models only apply to this particular content domain. As with all skills, it is important to provide students practice with modeling throughout their degree program to reinforce these skills. It is not necessary for students to engage in the full mathematical modeling process from conception to dissemination (Figure 2) in every course. For example, it may not make sense to have students engage in mathematical modeling activities such as deriving the Hardy-Weinberg mathematical equation in introductory biology. In contrast, it may be an explicit goal of an upper division genetics or evolution course to not only do so, but to revise the equation for absolute selection against a homozygous recessive genotype or to move from a one locus-two allele model to more loci or alleles or both. These goals can be supported by using non-mathematical modeling language in introductory biology through discussions of how to measure evolution and emphasizing the assumptions of the HWE model.

Another reason to engage students in mathematical modeling is that students come to biology with quantitative knowledge and reasoning skills (AP Calculus or Statistics, college-level math courses; see Jungck, 2011), but need practice retrieving and applying these skills properly in novel contexts (Hester et al., 2014). Students have exhibited significant learning gains when applying their quantitative skills to biological problems in mathematics courses designed for biology majors (Eaton and Highlander, 2017), and we encourage biology faculty to provide opportunities for students to practice and apply these

important skills in biology courses as well. If foundational mathematics and modeling knowledge is not practiced throughout the curriculum, proficiency in the skill will be lost.

## Additional Resources for Peeking into the Black Box with Students

One of our main goals in this essay is to inspire you to use models and modeling more frequently and emphasize model-based reasoning in your biology courses. Use of a variety of models can help students understand biological concepts and processes better and lay the foundation for incorporating more quantitative modeling into your courses. We invite you to join our community of faculty interested in using models in biology education. There are many resources available to support your efforts in this area. For mathematical modeling, we encourage you to read the GAIMME report (Garfunkel and Montgomery, 2016), explore the resources provided by the Society of Industrial and Applied Mathematics (SIAM, 2012; SIAM, 2014; SIAM, 2017), and join the QUBES community (<https://qubeshub.org>; Donovan et al., 2015), which provides resources, tools, and professional development opportunities around quantitative biology, especially through the Modeling Hub (<https://mmhub.qubeshub.org>, QUBES, 2017a). While we have not addressed the assessment of modeling skills in this essay, sample rubrics are available in the GAIMME report (Garfunkel and Montgomery, 2016; Bryce et al., 2016). We hope you will join us in helping students to peer inside the black box of modeling!

## Acknowledgments

This work was conducted as part of the “Unpacking the Black Box: Teaching Quantitative Biology” Working Group at the National Institute for Mathematical and Biological Synthesis (NIMBioS), sponsored by the National Science Foundation through NSF Award #DBI-1300426, with additional support from The University of Tennessee, Knoxville. Co-authors Glenn Ledder and Richard C. Schugart were supported as NIMBioS Sabbatical Fellows. We want to acknowledge the intellectual contributions of working group members who were not able to work on this paper including Gregory D. Goins, Ben G.

468 Fitzpatrick, and Edward F. (Joe) Redish and conversations with colleagues beyond the working group,  
469 especially Arietta E. Fleming-Davies and Anton E. Weisstein who gave valuable feedback on an earlier  
470 draft of this essay. Many thanks to our brave students and colleagues who allowed us to refine our ideas  
471 with them. Organization of this co-authorship and working group between meetings would not have been  
472 possible without the QUBES Hub supported by NSF Awards #DBI 1346584, DUE 1446269, DUE  
473 1446258, and DUE 1446284.

474

## References

- AAMC-HHMI Committee. (2009). *Scientific foundations for future physicians*. Washington, DC: Association of American Medical Colleges.
- Ainsworth, S. (1999). The functions of multiple representations. *Computers & Education*, 33, 131–152. [http://doi.org/10.1016/S0360-1315\(99\)00029-9](http://doi.org/10.1016/S0360-1315(99)00029-9)
- Allen, L. J., Brauer, F., Van den Driessche, P., & Wu, J. (2008). *Mathematical Epidemiology (Volume 1945)*. Berlin: Springer.
- American Association for the Advancement of Science (AAAS). (2011). *Vision and Change in Undergraduate Biology Education: A Call to Action*. Washington, DC: American Association for the Advancement of Science.
- Ashkenazy, H., Abadi, S., Martz, E., Chay, O., Mayrose, I., Pupko, T., & Ben-Tal, N. (2016). ConSurf 2016: an improved methodology to estimate and visualize evolutionary conservation in macromolecules. *Nucleic acids research*, 44(W1), W344–W350. <http://doi.org/10.1093/nar/gkw408>
- Barab, S. A., Hay, K. E., Barnett, M., & Keating, T. (2000). Virtual solar system project: Building understanding through model building. *Journal of Research in Science Teaching*, 37(7), 719–756. [http://doi.org/10.1002/1098-2736\(200009\)37:7<719::AID-TEA6>3.0.CO;2-V](http://doi.org/10.1002/1098-2736(200009)37:7<719::AID-TEA6>3.0.CO;2-V)
- Barton, C. K. (2000). Population genetics and evolution: a simulation exercise. In: *Tested Studies for Laboratory Teaching*, Vol. 21. Proceedings of the 21st Workshop/Conference of the Association for Biology Laboratory Education, ed. S. J. Karcher, 113–134. Retrieved April 30, 2018, from <http://www.ableweb.org/volumes/vol-21/6-barton.pdf>
- Biological ESTEEM Project. (2015a). DeFinetti 1.0. Retrieved August 24, 2017, from [http://bioquest.org/esteem/esteem\\_details.php?product\\_id=204](http://bioquest.org/esteem/esteem_details.php?product_id=204)
- Biological ESTEEM Project. (2015b). Deme 2.0. Retrieved August 24, 2017, from [http://bioquest.org/esteem/esteem\\_details.php?product\\_id=193](http://bioquest.org/esteem/esteem_details.php?product_id=193)
- Biological ESTEEM Project. (2015c). Evolution Through Natural Selection. Retrieved August 24, 2017, from [http://bioquest.org/esteem/esteem\\_details.php?product\\_id=7080](http://bioquest.org/esteem/esteem_details.php?product_id=7080)
- Biological ESTEEM Project. (2015d). An Introduction to ESTEEM. Retrieved August 24, 2017, from [http://bioquest.org/esteem/Intro\\_to\\_ESTEEM.pdf](http://bioquest.org/esteem/Intro_to_ESTEEM.pdf)
- Box, G. E. (1976). Science and statistics. *Journal of the American Statistical Association*, 71(356), 791–799. <http://doi.org/10.1080/01621459.1976.10480949>
- Brewe, E. (2008). Modeling theory applied: Modeling Instruction in introductory physics. *American Journal of Physics*, 76(12), 1155–1160. <http://doi.org/10.1119/1.2983148>
- Bryce, C. M., Baliga, V. B., De Nesnera, K. L., Fiack, D., Goetz, K., Tarjan, L. M., ... Gilbert, G. S. (2016). Exploring models in the biology classroom. *American Biology Teacher*, 8(1), 35–42. <http://doi.org/10.1525/abt.2016.78.1.35>

- College Board. (2015). AP Biology Course and Exam Description. Retrieved July 7, 2017, from <https://apcentral.collegeboard.org/pdf/ap-biology-course-and-exam-description.pdf?course=ap-biology>
- Cooper, A. K., & Oliver-Hoyo, M. T. (2017). Creating 3D physical models to probe student understanding of macromolecular structure. *Biochemistry and Molecular Biology Education*, 1–10. <http://doi.org/10.1002/bmb.21076>
- Cordell, S., Goldstein, G., Mueller-Dombois, D., Webb, D., & Vitousek, P. M. (1998). Physiological and morphological variation in *Metrosideros polymorpha*, a dominant Hawaiian tree species, along an altitudinal gradient: The role of phenotypic plasticity. *Oecologia*, 113(2), 188–196. <http://doi.org/10.1007/s004420050367>
- Dauer, J. T., Momsen, J. L., Speth, E. B., Makohon-Moore, S. C., & Long, T. M. (2013). Analyzing change in students' gene-to-evolution models in college-level introductory biology. *Journal of Research in Science Teaching*, 50(6), 639–659. <http://doi.org/10.1002/tea.21094>
- diSessa, A. (2004). Metarepresentation: Native competence and targets for instruction. *Cognition and Instruction*, 22(3), 293–331. [http://doi.org/10.1207/s1532690xci2203\\_2](http://doi.org/10.1207/s1532690xci2203_2)
- Dohn, N. B., Madsen, P. T., & Malte, H. (2009). The situational interest of undergraduate students in zoophysiology. *Advances in physiology education*, 33(3), 196–201. <http://doi.org/10.1152/advan.00038.2009>
- Donovan, S., Eaton, C. D., Gower, S. T., Jenkins, K. P., LaMar, M. D., Poli, D., ... Wojdak, J. M. (2015). QUBES: A community focused on supporting teaching and learning in quantitative biology. *Letters in Biomathematics*, 2(1), 46–55. <http://doi.org/10.1080/23737867.2015.1049969>
- Drott, M.T., Sarvary, M.A. 2016. Why did the snake cross the road? A population genetics and habitat conservation case study. National Center for Case Study Teaching in Science, University at Buffalo, State University of New York. Retrieved April 24, 2018, from [http://sciencecases.lib.buffalo.edu/cs/collection/detail.asp?case\\_id=884&id=884](http://sciencecases.lib.buffalo.edu/cs/collection/detail.asp?case_id=884&id=884)
- Eaton, C. D., Highlander, H. C., Dahlquist, K. D., LaMar, M. D., Ledder, G., & Schugart, R. C. (2018). A “Rule of Five” Framework for Models and Modeling to Unify Mathematicians and Biologists and Improve Student Learning. Revised version submitted to the journal *PRIMUS: Problems, Resources, and Issues in Mathematics Undergraduate Studies* on April 28, 2017, available from the *arXiv* preprint server, <https://arxiv.org/abs/1607.02165v2>
- Eaton, C. D., & Highlander, H. C. (2017). The case for biocalculus: Design, retention, and student performance. *CBE Life Sciences Education*, 16(2), 1–11. <http://doi.org/10.1187/cbe.15-04-0096>
- Feser, J., Vasaly, H., & Herrera, J. (2013). On the edge of mathematics and biology integration: improving quantitative skills in undergraduate biology education. *CBE-Life Sciences Education*, 12(2), 124–128. <http://doi.org/10.1187/cbe.13-03-0057>
- FirstGlance Development Team (2018) FirstGlance in Jmol. Retrieved from <http://firstglance.jmol.org> 8 May 2018

- 547 Fretz, E. B., Wu, H. H.-K., Zhang, B., Davis, E. A., Krajcik, J. S., & Soloway, E. (2002). An  
 548 investigation of software scaffolds supporting modeling practices. *Research in Science Education*,  
 549 32(4), 567–589. <http://doi.org/10.1023/A:1022400817926>
- 550 Frigg, R. (2002) Models and representation: why structures are not enough. Measurement in physics and  
 551 economics discussion papers, Dietsch, Peter (ed.) 25/02. The London School of Economics and  
 552 Political Science, London, UK. Retrieved May 8, 2018, from  
 553 [http://www.romanfrigg.org/writings/Models\\_and\\_Representation.pdf](http://www.romanfrigg.org/writings/Models_and_Representation.pdf)
- 554 Frigg, R., Hartmann, S. (2017) "Models in Science", *The Stanford Encyclopedia of Philosophy (Spring*  
 555 *2017 Edition)*, Edward N. Zalta (ed.), URL  
 556 <<https://plato.stanford.edu/archives/spr2017/entries/models-science/>>
- 557 Froehlich, J. W., and London, M. R. (1996). Beanbag Population Genetics. *AnthroNotes*, vol. 18(2).  
 558 Retrieved April 30, 2018, from  
 559 [https://anthropology.si.edu/outreach/Teaching\\_Activities/pdf/BEANBAG.96.pdf](https://anthropology.si.edu/outreach/Teaching_Activities/pdf/BEANBAG.96.pdf)
- 560 Garfunkel, S., & Montgomery, M. (Eds.). (2016). *GAIMME Report: Guideline for Assessment and*  
 561 *Instruction in Mathematical Modeling Education*. Bedford, MA: Consortium for Mathematics and Its  
 562 Applications. ISBN: 978-1-611974-43-0
- 563 Gilbert, J. K. (2004). Models and modelling: Routes to more authentic science education. *International*  
 564 *Journal of Science and Mathematics Education*, 2, 115–130. [http://doi.org/10.1007/s10763-004-](http://doi.org/10.1007/s10763-004-3186-4)  
 565 3186-4
- 566 Gilbert, S. W. (1991). Model building and a definition of science. *Journal of Research in Science*  
 567 *Teaching*, 28(1), 73–79. <http://doi.org/10.1002/tea.3660280107>
- 568 Gillooly, J. F., Gomez, J. P., Mavrodiev, E. V., Rong, Y., & McLamore, E. S. (2016). Body mass scaling  
 569 of passive oxygen diffusion in endotherms and ectotherms. *Proceedings of the National Academy of*  
 570 *Sciences U S A*, 113(19), 5340-5345. <http://doi.org/10.1073/pnas.1519617113>
- 571 Goel, A. K., & Stroulia, E. (1996). Functional device models and model-based diagnosis in adaptive  
 572 design. *AI EDAM*, 10(4), 355-370. <https://doi.org/10.1017/S0890060400001670>
- 573 Gouvea, J., & Passmore, C. (2017). ‘Models of’ versus ‘Models for’: Toward an Agent-Based Conception  
 574 of Modeling in the Science Classroom. *Science & Education*, 26(1-2), 49-63.  
 575 <http://doi.org/10.1007/s11191-017-9884-4>
- 576 Grimm, V., Augusiak, J., Focks, A., Frank, B. M., Gabsi, F., Johnston, A. S. A., ... Railsback, S. F.  
 577 (2014). Towards better modelling and decision support: Documenting model development, testing,  
 578 and analysis using TRACE. *Ecological Modelling*, 280, 129–139.  
 579 <http://doi.org/10.1016/j.ecolmodel.2014.01.018>
- 580 Hardy, G. H. (1908). Mendelian Proportions in a Mixed Population. *Science*, 28(706), 49–50.  
 581 <http://doi.org/10.1126/science.28.706.49>
- 582 Hestenes, D. (1992). Modeling games in the Newtonian world. *American Journal of Physics*, 60(8), 732-  
 583 748. <http://doi.org/10.1119/1.17080>

- 584 Hester, S., Buxner, S., Elfring, L., & Nagy, L. (2014). Integrating quantitative thinking into an  
585 introductory biology course improves students' mathematical reasoning in biological contexts. *CBE -*  
586 *Life Sciences Education*, 13, 54–64. <http://doi.org/10.1187/cbe.13-07-0129>
- 587 Hidi, S., & Harackiewicz, J. M. (2000). Motivating the academically unmotivated: A critical issue for the  
588 21st century. *Review of educational research*, 70(2), 151-179.  
589 <http://doi.org/10.3102/00346543070002151>
- 590 Hmelo-Silver, C. E., Jordan, R., Eberbach, C., & Sinha, S. (2017). Systems learning with a conceptual  
591 representation: A quasi-experimental study. *Instructional Science*, 45(1), 53-72.  
592 <http://doi.org/10.1007/s11251-016-9392-y>
- 593 Hoefnagel, M. H., Starrenburg, M. J., Martens, D. E., Hugenholtz, J., Kleerebezem, M., Van Swam, I. I.,  
594 ... & Snoep, J. L. (2002). Metabolic engineering of lactic acid bacteria, the combined approach:  
595 kinetic modelling, metabolic control and experimental analysis. *Microbiology*, 148(4), 1003-1013.  
596 <http://doi.org/10.1099/00221287-148-4-1003>
- 597 Hogan, K., & Thomas, D. (2001). Cognitive comparisons of students' systems modeling in ecology.  
598 *Journal of Science Education and Technology*, 10(4), 319-345. <http://doi.org/10.1023/A:1012243102>
- 599 Ifenthaler, D. (2011). Identifying cross-domain distinguishing features of cognitive structure. *Educational*  
600 *Technology Research and Development*, 59(6), 817-840. <http://doi.org/10.1007/s11423-011-9207-4>
- 601 Jaswal, S. S., O'Hara, P. B., Williamson, P. L., & Springer, A. L. (2013). Teaching structure: Student use  
602 of software tools for understanding macromolecular structure in an undergraduate biochemistry  
603 course. *Biochemistry and Molecular Biology Education*, 41(5), 351-359.  
604 <http://doi.org/10.1002/bmb.20718>
- 605 Jungck, J. R. (1985). A problem posing approach to biology education. *The American Biology Teacher*,  
606 47(5), 264-266. <http://doi.org/10.2307/4448046>
- 607 Jungck, J. R. (2011). Mathematical biology education: modeling makes meaning. *Mathematical*  
608 *Modelling of Natural Phenomena*, 6(6), 1-21. <http://doi.org/10.1051/mmnp/20116601>
- 609 Jungck, J. R., Gaff, H., & Weisstein, A. E. (2010). Mathematical manipulative models: In defense of  
610 “Beanbag Biology”. *CBE-Life Sciences Education*, 9(3), 201-211. [http://doi.org/10.1187/cbe.10-03-](http://doi.org/10.1187/cbe.10-03-0040)  
611 0040
- 612 Just, W., Callender, H., LaMar, M. D., & Toporikova, N. (2015). Transmission of infectious diseases:  
613 Data, models, and simulations. In R. Robeva (Ed.), *Algebraic and Discrete Mathematical Methods*  
614 *for Modern Biology* (pp. 193–215). San Diego, CA: Academic Press.
- 615 Kain, K. H. (2009). Telomeres and tetrahymena: an interview with Elizabeth Blackburn. *Disease models*  
616 *& mechanisms*, 2(11-12), 534-537. <http://doi.org/10.1242/dmm.003418>
- 617 Keeling, M.J., Danon, L. (2009) Mathematical modelling of infectious diseases, *British Medical Bulletin*,  
618 92(1): 33–42, <https://doi.org/10.1093/bmb/ldp038>



- 619 Kermack, W. and McKendrick, A. (1927). A contribution to the mathematical theory of epidemics.  
620 *Proceedings of the Royal Society 115A*:700–721. <http://doi.org/10.1098/rspa.1927.0118>
- 621 Kline, M. (1980). *Mathematics: The Loss of Certainty*. New York, NY. ISBN: 978-0195030853
- 622 Kutmon, M., Riutta, A., Nunes, N., Hanspers, K., Willighagen, E. L., Bohler, A., ... & Coort, S. L.  
623 (2015). WikiPathways: capturing the full diversity of pathway knowledge. *Nucleic acids research*,  
624 *44*(D1), D488-D494. <http://doi.org/10.1093/nar/gkv1024>
- 625 Lander, A. D. (2010). The edges of understanding. *BMC Biology*, 8, 40. [http://doi.org/10.1186/1741-](http://doi.org/10.1186/1741-7007-8-40)  
626 7007-8-40
- 627 Latour, B. (1999). *Pandora's Hope: On the Reality of Science Studies*, Cambridge, MA: Harvard  
628 University Press, xi + 324 pp. ISBN: 0674653351
- 629 Lehrer, R., & Schauble, L. (2005). Developing modeling and argument in the elementary grades. In  
630 Romberg, T. A., Carpenter, T. P., & Dremock, F. (Eds.). *Understanding mathematics and science*  
631 *matters*. Routledge, 29–53.
- 632 Lehrer, R., & Schauble, L. (2010). What kind of explanation is a model? In M. K. Stein (Ed.),  
633 *Instructional explanations in the disciplines* (pp. 9–22). New York: Springer.
- 634 Long, T. M., Dauer, J. T., Kostelnik, K. M., Momsen, J. L., Wyse, S. A., Speth, E. B., & Ebert-May, D.  
635 (2014). Fostering ecoliteracy through model-based instruction. *Frontiers in Ecology and the*  
636 *Environment*, 12(2), 138–139. <http://doi.org/10.1890/1540-9295-12.2.138>
- 637 Louca, L. T., & Zacharia, Z. C. (2012). Modeling-based learning in science education: cognitive,  
638 metacognitive, social, material and epistemological contributions. *Educational Review*, 64(4), 471-  
639 492. <http://doi.org/10.1080/00131911.2011.628748>
- 640 Macal, C. M., & North, M. J. (2006). Tutorial on agent-based modeling and simulation part 2: How to  
641 model with agents. In L. F. Perrone, F. P. Wieland, J. Liu, B. J. Lawson, D. M. Nicol, & R. M.  
642 Fujimoto (Eds.), *Proceedings of the Winter 2006 Simulation Conference* (pp. 73–83).
- 643 Marbach-Ad, G., & Stavy, R. (2000). Students' cellular and molecular explanations of genetic  
644 phenomena. *Journal of Biological Education*, 34(4), 200-205.  
645 <http://doi.org/10.1080/00219266.2000.9655718>
- 646 Mayes, R., Forrester, J., Christus, J. Peterson, F. Bonilla, R. and Yestness, N. (2013a). Quantitative  
647 reasoning in environmental science: A learning progression. *International Journal of Science*  
648 *Education*, 36, (4) 635-658. [new one from Bob]
- 649 Mayes, R. L., Peterson, F., & Bonilla, R. (2013b). Quantitative Reasoning Learning Progressions for  
650 Environmental Science: Developing a Framework. *Numeracy*, 6(1), Article 4.  
651 <http://doi.org/10.5038/1936-4660.6.1.4>
- 652 Moulton J, Fidelis K, Kryshtafovych A, Schwede T, Tramontano A. Critical assessment of methods of  
653 protein structure prediction (CASP)—Round XII. *Proteins*. 2018;86:7–15.  
654 <http://doi.org/10.1002/prot.25415>

- 655 Mulder, Y. G., Bollen, L., de Jong, T., & Lazonder, A. W. (2016). Scaffolding learning by modelling:  
656 The effects of partially worked-out models. *Journal of research in science teaching*, 53(3), 502-523.  
657 <http://doi.org/10.1002/tea.21260>
- 658 Nersessian, N. J. (2002). The cognitive basis of model-based reasoning in science. Pages 133-153 in P.  
659 Carruthers, S. Stich, and M. Siegal, editors. *The cognitive basis of science*. Cambridge University  
660 Press, Cambridge, UK. <http://doi.org/10.1017/CBO9780511613517.008>
- 661 Nersessian NJ. (1999). Model-based reasoning in conceptual change. In *Model-Based Reasoning in*  
662 *Scientific Discovery*, ed. L Magnani, NJ Nersessian, P Thagard, pp. 5–22. New York: Kluwer  
663 Acad./Plenum. [http://doi.org/10.1007/978-1-4615-4813-3\\_1](http://doi.org/10.1007/978-1-4615-4813-3_1)
- 664 NGSS Lead States. (2013). *Next Generation Science Standards: For States, By States*. Washington, DC:  
665 National Academies Press.
- 666 Odenbaugh, J. A. Y. (2005). Idealized , inaccurate but successful : A pragmatic approach to evaluating  
667 models in theoretical ecology. *Biology and Philosophy*, 20, 231–255. [http://doi.org/10.1007/s10539-](http://doi.org/10.1007/s10539-004-0478-6)  
668 004-0478-6
- 669 Papaevripidou, M., Constantinou, C. P., & Zacharia, Z. C. (2007). Modeling complex marine ecosystems:  
670 An investigation of two teaching approaches with fifth graders. *Journal of Computer Assisted*  
671 *Learning*, 23(2), 145–157. <http://doi.org/10.1111/j.1365-2729.2006.00217.x>
- 672 Papaevripidou, M., & Zacharia, Z. C. (2015). Examining how students' knowledge of the subject domain  
673 affects their process of modeling in a computer programming environment. *Journal of Computers in*  
674 *Education*, 2(3), 251–282. <http://doi.org/10.1007/s40692-015-0034-1>
- 675 Pearsall, N. R., Skipper, J. O. E. L. J., & Mintzes, J. J. (1997). Knowledge restructuring in the life  
676 sciences: A longitudinal study of conceptual change in biology. *Science Education*, 81(2), 193–215.  
677 [http://doi.org/10.1002/\(SICI\)1098-237X\(199704\)81:2<193::AID-SCE5>3.0.CO;2-A](http://doi.org/10.1002/(SICI)1098-237X(199704)81:2<193::AID-SCE5>3.0.CO;2-A)
- 678 Peterson, N. S., & Jungck, J. R. (1988). Problem-posing, problem-solving and persuasion in biology  
679 education. *Academic Computing*, 2(6), 14–17 & 48–50.
- 680 QUBES. (2017a). Math Modeling Hub. Retrieved August 24, 2017, from <https://mmhub.qubeshub.org>
- 681 QUBES. (2017b). QUBES – Home. Retrieved August 24, 2017, from <https://qubeshub.org/>
- 682 Railsback, S. F., & Grimm, V. (2011). *Agent-based and Individual-based Modeling: A Practical*  
683 *Introduction*. Princeton, NJ: Princeton University Press.
- 684 Rothstein, D., & Santana, L. (2011). Teaching students to ask their own questions. *Harvard Education*  
685 *Letter*, 27(5), 1–2.
- 686 Schuchardt, A. M., & Schunn, C. D. (2016). Modeling scientific processes with mathematics equations  
687 enhances student qualitative conceptual understanding and quantitative problem solving. *Science*  
688 *Education*, 100(2), 290-320. <http://doi.org/10.1002/sce.21198>
- 689 Schwarz, C. V., Reiser, B. J., Davis, E. A., Kenyon, L., Fortus, D., Shwartz, Y., ... Ache, A. (2009).  
690 Developing a learning progression for scientific modeling: Making scientific modeling accessible and

- 691 meaningful for learners. *Journal of Research in Science Teaching*, 46(6), 632–654.  
692 <http://doi.org/10.1002/tea.20311>
- 693 Schwarz, C. V., & White, B. Y. (2005). Metamodeling knowledge: Developing students' understanding  
694 of scientific modeling. *Cognition and instruction*, 23(2), 165-205.  
695 [http://doi.org/10.1207/s1532690xci2302\\_1](http://doi.org/10.1207/s1532690xci2302_1)
- 696 Servedio, M. R., Brandvain, Y., Dhole, S., Fitzpatrick, C. L., Goldberg, E. E., Stern, C. A., Van Cleve, J.  
697 and Yeh, D. J., (2014). Not just a theory—the utility of mathematical models in evolutionary  
698 biology. *PLoS biology*, 12(12), e1002017. <http://doi.org/10.1371/journal.pbio.1002017>
- 699 Sheehy, R. (2017) Population genetics simulation program. Retrieved August 24, 2017, from  
700 [http://www.radford.edu/~rsheehy/Gen\\_flash/popgen/](http://www.radford.edu/~rsheehy/Gen_flash/popgen/)
- 701 Sim, J. H., & Daniel, E. G. S. (2014). Representational competence in chemistry: A comparison between  
702 students with different levels of understanding of basic chemical concepts and chemical  
703 representations. *Cogent Education*, 1(1), 991180. <http://doi.org/10.1080/2331186X.2014.991180>
- 704 Simundza, G. (2006). The fifth rule: direct experience of mathematics. *MAA NOTES*, 69, 320.
- 705 Sins, P. H., Savelsbergh, E. R., & van Joolingen, W. R. (2005). The Difficult Process of Scientific  
706 Modelling: An analysis of novices' reasoning during computer-based modelling. *International*  
707 *Journal of Science Education*, 27(14), 1695-1721. <http://doi.org/10.1080/09500690500206408>
- 708 Society for Industrial and Applied Mathematics (SIAM). (2017). Math Modeling Challenge Resources.  
709 Retrieved August 24, 2017, from <https://m3challenge.siam.org/resources>
- 710 Society for Industrial and Applied Mathematics (SIAM). (2012). *Modeling across the curriculum I*.  
711 Retrieved August 24, 2017, from [http://www.siam.org/reports/modeling\\_12.pdf](http://www.siam.org/reports/modeling_12.pdf)
- 712 Society for Industrial and Applied Mathematics (SIAM). (2014). *Modeling across the curriculum II*.  
713 Retrieved August 24, 2017, from [http://www.siam.org/reports/ModelingAcross%20Curr\\_2014.pdf](http://www.siam.org/reports/ModelingAcross%20Curr_2014.pdf)
- 714 Soderberg, P., & Price, F. (2003). An examination of problem-based teaching and learning in population  
715 genetics and evolution using EVOLVE, a computer simulation. *International Journal of Science*  
716 *Education*, 25(1), 35-55. <http://doi.org/10.1080/09500690110095285>
- 717 Speth, E. B., Shaw, N., Momsen, J., Reinagel, A., Le, P., Taqieddin, R., & Long, T. (2014). Introductory  
718 biology students' conceptual models and explanations of the origin of variation. *CBE Life Sciences*  
719 *Education*, 13(3), 529–539. <http://doi.org/10.1187/cbe.14-02-0020>
- 720 Svoboda, J., & Passmore, C. (2011). The Strategies of Modeling in Biology Education. *Science &*  
721 *Education*, 22, 119–142. <http://doi.org/10.1007/s11191-011-9425-5>
- 722 Thompson, P. W. (2011). Quantitative reasoning and mathematical modeling. In L. L. Hatfield, S.  
723 Chamberlain, & S. Belbase (Eds.), *New Perspectives and Directions for Collaborative Research in*  
724 *Mathematics Education*. Laramie, WY: University of Wyoming.
- 725 Tomasi, J. (1988). Models and modeling in theoretical chemistry. *Journal of Molecular Structure:*  
726 *THEOCHEM*, 179(1), 273–292. [http://doi.org/10.1016/0166-1280\(88\)80128-3](http://doi.org/10.1016/0166-1280(88)80128-3)

- 727 Understanding Science. (2016). The Real Process of Science. Retrieved August 24, 2017, from  
728 [http://undsci.berkeley.edu/article/howscienceworks\\_02](http://undsci.berkeley.edu/article/howscienceworks_02)
- 729 Weisstein, A. E. (2011). Building mathematical models and biological insight in an introductory biology  
730 course. *Mathematical Modelling of Natural Phenomena*, 6(6), 198-214.  
731 <http://doi.org/10.1051/mmnp/20116610>
- 732 WikiPathways. (2017) WikiPathways. Retrieved August 24, 2017, from <http://wikipathways.org>
- 733 Wilensky, U., & Reisman, K. (2006). Thinking like a wolf, a sheep, or a firefly: Learning biology through  
734 constructing and testing computational theories—an embodied modeling approach. *Cognition and*  
735 *instruction*, 24(2), 171-209. [http://doi.org/10.1207/s1532690xci2402\\_1](http://doi.org/10.1207/s1532690xci2402_1)
- 736 Wimsatt, William. (1987), “False Models as Means to Truer Theories”, in N. Nitecki and A.  
737 Hoffman(eds.), *Neutral Models in Biology*, Oxford: Oxford University Press, 23–55. ISBN:  
738 9780195050998
- 739 Windschitl, M., Thompson, J., & Braaten, M. (2008). Beyond the scientific method: Model-based inquiry  
740 as a new paradigm of preference for school science investigations. *Science Education*, 92, 941–967.  
741 <http://doi.org/10.1002/sce.20259>

## Figure Legends

**Figure 1.** Different model representations. (A) a physical DNA model put together by students (see also, Cooper & Oliver-Hoyo, 2017) and a screenshot of the PopGen simulator (Sheehy, 2017) are examples of experiential model representations; (B) a schematic of a gene is an example of a visual model representation; (C) a statement of the null hypothesis of the Hardy-Weinberg model is an example of a verbal model representation; (D) a graph and data table of genotype frequencies are examples of numerical model representations; (E) the Hardy-Weinberg equations are an example of a symbolic model representation.

**Figure 2.** The parallel and iterative nature of the process of mathematical modeling and the process of experimental science as instances of model-based reasoning (cf. Servedio, 2014). When approaching a new problem, one begins by identifying the problem and organizing contextual information before proceeding to asking questions and formulating hypotheses, which then inform the seeking of evidence. Evidence can be obtained through either mathematical models or experiments (shown), or other avenues such as observational field studies, evolutionary reconstruction approaches, or meta-analyses. In each track (left and right), the steps of the mathematical and experimental model design have clear parallels. Validation (dashed arrow) occurs when experimental data are compared to model output or vice versa. The analysis of mathematical model and experimental model results are used to construct an argument for a particular biological interpretation, which is documented (Grimm et al., 2014) and disseminated to other scientists. This, in most cases, leads to even more questions. Each arrow is a modeling activity that can be performed with students. While this diagram was drawn in a top-to-bottom, linear fashion to facilitate easy viewing, we recognize that actual practice may be messier, requiring entering the diagram at different points, traversing the steps in a different order, and repeating steps (Eaton et al., 2018; Understanding Science)<sup>1</sup>

## Footnote

<sup>1</sup> Interestingly, this figure, depicting the relationships and parallels between mathematical and experimental approaches to modeling, is itself a model. It has served an important purpose in the negotiation of our shared understanding of modeling over the course of collaboratively writing this paper. We have actively used this model as a point of focus during our attempts to articulate our claims about the modeling process. Questions like, “What exactly does this box represent?”, “Why is this word used instead of another?”, and “Why are there unidirectional arrows here but bidirectional arrows in another place?” have been asked by the biologists and mathematicians to each other when developing our ideas. This figure has been refined many times.

**Box 1: Definitions of terms**

**Model:** A simplified representation of real-world objects and their mechanistic or functional relationships, constructed for a purpose (Frigg, 2002; Lehrer and Schauble, 2010; Eaton et al., 2018)

**Model utility:** The purpose(s) for constructing and using the model, *e.g.* developing conceptual frameworks or making accurate predictions (see Table 1; Odenbaugh, 2005).

**Modeling:** An iterative process in which a model is proposed, explored, validated, and refined (the arrows of Figure 2; Louca and Zacharia, 2012; Garfunkel and Montgomery, 2016).

**Mathematization:** The modeling process of going from a visual schematic or verbal description of the model and assumptions to a symbolic mathematical model representation.

**Model exploration:** Depending on the type of mathematical model, model exploration can consist of mathematical analyses or computer simulations to observe the behavior of the model as a function of its assumptions, inputs, and parameters.

**Model validation:** The process of assessing a model's output and assumptions with regard to its desired utility (is it addressing our goals?) and accuracy (is it consistent with other lines of evidence, *e.g.*, experimental data, observations and/or different models?).

**Model refinement:** Modifying aspects of the model, including changing the objects, processes and/or relationships.

**Reframe:** Incorporating the model and results into the broader set of scientific work, leading to new questions, hypotheses, or foci for scientific exploration.

**Model-based Reasoning:** Forms of inquiry based on the process of modeling; using models to understand biological concepts.

**Box 2: Practicing model building and refinement in introductory biology**

Modeling in my introductory biology course is used by students to organize their thoughts about the components of a system in order to better understand the system dynamics. There are some basic tenets to modeling in my class that are common to all the contexts. I primarily use visual models based loosely on the Structure, Behavior, Function (SBF) framework (Goel et al., 1996; Hmelo-Silver et al., 2017). These models have labeled boxes that represent structures, usually nouns, connected by labeled arrows that represent behaviors, usually verbs or modifiers, created to represent the function of a system. While the behaviors can be mathematical, in introductory biology I usually keep these non-mathematical. I have students make sure all the boxes, and especially all the arrows, are labeled in the model to tell a single consistent story. Here are two instances where I use this type of modeling in the classroom.

1. Hox genes

Hox genes, the homeotic box genes found in all animals, produce regulatory transcription factors that turn on their target genes during development. To create a systems model for Hox genes, students are provided with a list of components and the list of relationships in order to show how Hox genes usually activate multiple target genes when they bind to enhancer regions on chromosomes. The objectives are to reinforce the mechanistic relationship of gene to protein and establish a baseline model that will be used in the following assignment to explore the evolution of development. With this model successfully built, students are then asked to modify the model to predict what would happen due to mutations in different regions of the genome, considering the proximate and ultimate impacts of these mutations on animal phenotype. This model helps lay the groundwork for a discussion of biodiversity.

2. Tracing matter and energy in animals and plants

When learning about animal and plant systems, such as the circulatory system or photosynthesis, it is easy for students to focus on the components, while missing the connections (i.e., the inputs and outputs) to other systems within the organism. The purpose of these models is for students to develop a conceptual framework that shows conservation of matter in order to describe how any cell in the organism obtains the necessary nutrients for biosynthesis and cellular respiration and eliminates wastes. Students trace carbon, hydrogen, and oxygen atoms through the organism, incorporating a vast amount of knowledge about different systems. This activity reveals both the simplicity and complexity of organ systems and gives students the space to explore their own knowledge about system interconnectivity.



**Box 3:** Using mathematical models to test hypotheses in introductory biology

The fields of ecology and evolutionary biology have a rich history of mathematical modeling. Therefore, in my Introductory Biology: Evolution, Biodiversity, and Ecology course, I stress the importance of models as tools for biologists. One of my course learning objectives is that students will be able to interpret the output of models to make claims about hypotheses. I have students complete several problem sets over the course of the semester in which they are expected to use a given mathematical model to test a hypothesis. I anchor all of my problem sets in a conservation or animal physiology context to promote interest, since my students are majoring in zoology, marine biology, environmental science, or biology with an emphasis in ecology. Here are two examples.

**1. Hardy-Weinberg Equilibrium Model**

I use a modified version of the NCCSTS “Why did the snake cross the road?” case study (Drott and Sarvary, 2016), which is anchored in the conservation context of the effect of a road bisecting the habitat of the timber rattlesnake. Students test the hypothesis that snakes living on either side of a road represent one large population versus two smaller populations using the Hardy-Weinberg Equilibrium model. The goal is for students to practice calculating expected genotype frequencies, compare them to observed genotype frequencies, and draw conclusions about whether or not populations are in Hardy-Weinberg Equilibrium.

**2. Fick’s Law of Diffusion**

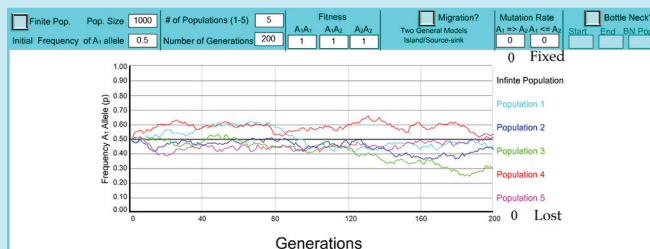
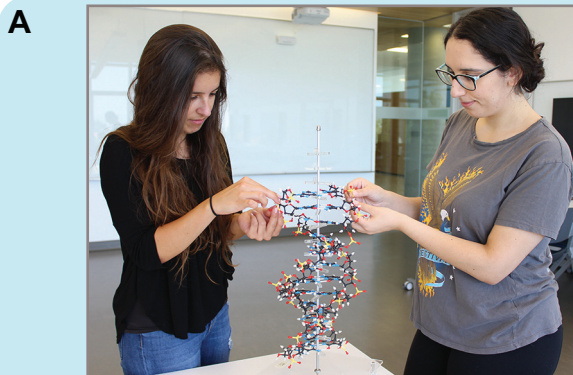
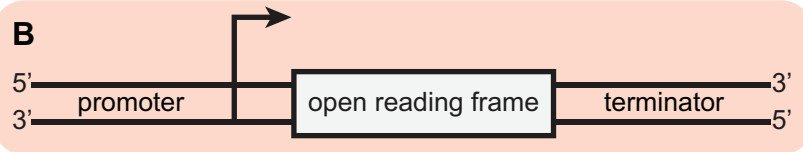
In one of the animal physiology problem sets, I ask students, what respiratory structural traits allow for greater oxygen uptake rates? I then present them with the mathematical model for Fick’s Law of Diffusion and have them generate a hypothesis based on the model. Students need to understand that because thickness of the respiration barrier is in the denominator of the equation, a decrease in respiration barrier thickness would be predicted to increase oxygen uptake rates, whereas an increase in respiration surface area would be predicted to increase oxygen uptake rates. I then give students data from a paper by Gillooly and colleagues (2016) that includes all of the parameters needed to calculate oxygen uptake rates for a variety of vertebrate animals using Fick’s Law of Diffusion. The students graph the oxygen uptake rates they calculated as a function of respiration surface area and barrier thickness. They then use the graphs to determine whether their hypotheses were supported or not.

1 **Table 1:** Five types of model utility as described by Odenbaugh (2005) with example models.

Model Utility	Example Model
Simple, unrealistic models for exploring complex systems	<ul style="list-style-type: none"> <li>• Using Tetrahymena is used as a model organism to study telomeres because it has tens of thousands of short linear chromosomes (Kain, 2009).</li> <li>• A model of an epidemic with different initial populations of susceptible, infected, and resistant individuals could be explored with different rules for transmission and recovery to provide insights into how different diseases spread through a population (Allen et al., 2008; Weisstein, 2011; Just et al., 2015).</li> </ul>
Exploring unknown possibilities	<ul style="list-style-type: none"> <li>• Building 3D models based on predicted protein structures could be used to understand drug-target interactions.</li> <li>• Agent-based models could be used to identify simple interaction rules that can lead to different emergent population level behaviors like flocking (Macal and North, 2006; Railsback and Grimm, 2011).</li> </ul>
Developing conceptual frameworks	<ul style="list-style-type: none"> <li>• A pathway diagram is a conceptual model summarizing experimental results (examples can be found at WikiPathways, <a href="http://wikipathways.org">http://wikipathways.org</a>; Kutmon et al., 2015).</li> <li>• The Hardy-Weinberg null model can provide a starting point for explaining diverse evolutionary forces.</li> </ul>
Making accurate predictions	<ul style="list-style-type: none"> <li>• Data-driven population models of fish stocks inform sustainable harvests.</li> <li>• An enzyme kinetic model of pyruvate carbon distribution in lactic acid bacteria accurately predicted which genes to manipulate to increase flavor compound production (Hoefnagel et al., 2002).</li> </ul>
Generating causal explanations	<ul style="list-style-type: none"> <li>• A common garden experiment was used to determine whether differences in traits among populations of a plant species is due to genetic differences or phenotypic plasticity (Cordell et al., 1998).</li> <li>• The Hodgkin-Huxley symbolic model of ion flow across cell membranes helps to explain the all or none firing of action potentials.</li> </ul>

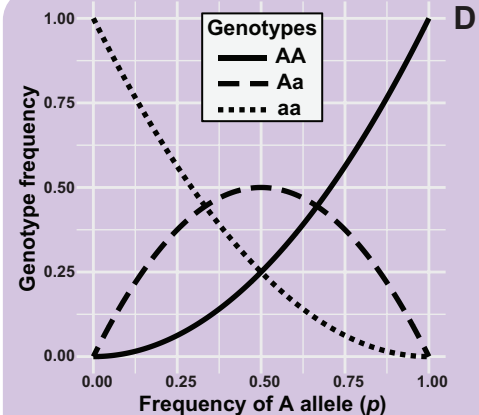
1 **Table 2:** Computational models for population genetics.

Computational Model	URL
PopGen (Sheehy, 2017)	<a href="http://www.radford.edu/~rsheehy/Gen_flash/popgen/">http://www.radford.edu/~rsheehy/Gen_flash/popgen/</a>
Deme 2.0 (Biological ESTEEM Project., 2015b)	<a href="http://bioquest.org/esteem/esteem_details.php?product_id=193">http://bioquest.org/esteem/esteem_details.php?product_id=193</a>
DeFinetti 1.0 (Biological ESTEEM Project., 2015a)	<a href="http://bioquest.org/esteem/esteem_details.php?product_id=204">http://bioquest.org/esteem/esteem_details.php?product_id=204</a>
Evolution Through Natural Selection (Biological ESTEEM Project., 2015c)	<a href="http://bioquest.org/esteem/esteem_details.php?product_id=7080">http://bioquest.org/esteem/esteem_details.php?product_id=7080</a>



**C**

*Allele and genotype frequencies are not changing from generation to generation in the absence of evolutionary forces.*



	<u>AA</u>	<u>Aa</u>	<u>aa</u>
F <sub>0</sub> :	49	111	56
F <sub>1</sub> :	51	109	53

**E**

$$p + q = 1$$

$$p^2 + 2pq + q^2 = 1$$

# Model-Based Reasoning

