STUDIA METODOLOGICZNE
NR 39 • 2019, 73–98
DOI: 10.14746/sm.2019.39.3

Steven Hecht Orzack, Brian McLoone

Modeling in Biology: looking backward and looking forward

Abstract. Understanding modeling in biology requires understanding how biology is organized as a discipline and how this organization influences the research practices of biologists. Biology includes a wide range of sub-disciplines, such as cell biology, population biology, evolutionary biology, molecular biology, and systems biology among others. Biologists in sub-disciplines such as cell, molecular, and systems biology believe that the use of a few experimental models allows them to discover biological universals, whereas biologists in sub-disciplines such as ecology and evolutionary biology believe that the use of many different experimental and mathematical models is necessary in order to do this. Many practitioners of both approaches misunderstand best practices of modeling, especially those related to model testing. We stress the need for biologists to better engage with best practices and for philosophers of biology providing normative guidance for biologists to better engage with current developments in biology. This is especially important as biology transitions from a "data-poor" to a "data-rich" discipline. If 21st century biology is going to capitalize on the unprecedented availability of ecological, evolutionary, and molecular data, of computational resources, and of mathematical and statistical tools, biologists will need a better understanding of what modeling is and can be.

Keywords: biology, model building, model testing, philosophy of biology, subdisciplines of biology.

"Biology" is the study of nature. It dates as far back as our ancestors paid attention to the benefits and hazards of their surroundings. A possible vestige of this study may be the common but not universal fear of snakes and spiders, which could be due to long-ago observations that some are dangerous (Rakison 2018).

More formal biology is less ancient but still has a long history. Aristotle was a biologist among other things and his contributions to the discipline are profound (Egerton 1975, 2001; Balme 1987; Gotthelf 1999, 2012; Leroi 2014). His study of nature was not just passive observation; for example, he may have sequestered fish in order to study their foraging behavior (Tipton 2008).

Today, biology is organized into a wide range of sub-disciplines such as cell biology, ecology, evolutionary biology, molecular biology, systems biology, and population biology. These have divisions as well, such as experimental population biology and theoretical population biology. There are also sub-disciplines that focus on particular kinds of organisms, such as botany, entomology, mammalogy, microbiology, ornithology, virology, and zoology. Use of these descriptors depends on context. A biologist might say to another biologist "My colleague is a microbiologist but I am an ecologist", whereas in conversation with the public (s)he might say "My colleague and I are biologists".

1. Sub-disciplinary differences in the way biologists create, understand, and use models

We now discuss sub-disciplinary differences in modeling practices and in the institutional context in which modeling occurs. In doing so, we use this "folk" definition enunciated by Barbour (1974, p. 6): "[a model is] a symbolic representation of selected aspects of behavior of a complex system for particular purposes…". We believe that this captures the loose definition of model used by many biologists (see also Lewontin 1968); Frigg and Hartmann (2012) review more nuanced philosophical considerations of what models are and how they are used in science.

Models in biology can be computational, experimental, mathematical, or verbal. Even within these categories, there is variety. For example, an experimental model can be an actual organism (e.g., Tickoo and Russell 2002) or a physical representation of an organism (e.g., Colbert 1962). Mathematical models can be deterministic or stochastic (e.g., Bartlett 1956).

Most biochemists, cell biologists, molecular biologists, neurologists, pharmacologists, and systems biologists use experimental models involving one

or a few types of organisms. A few of the almost innumerable examples of such "microcosms" as models are the use of cells in culture to investigate a cellular process (e.g., Szostak, Orr-Weaver, Rothstein, & Stahl, 1983) and in-vitro analysis of an enzyme and substrate to investigate the enzyme's in-vivo activity (e.g., Tcherkez et al., 2013; Torres, Mateo, Melendez-Hevia, & Kacser, 1986). Mathematical models are rare in these sub-disciplines although some are very influential (e.g., Hodgkin and Huxley 1952; Kacser and Burns 1973; Byrne 2010).

Models in these sub-disciplines are usually conceived of as providing insight into "the" biology of the cell, enzyme, or pathway studied. This typological conception is based upon beliefs that the model is "about" large classes of organisms and not specifically about the species used, that differences among species are "noise", and that related species would provide only redundant information. These beliefs sustain careers dedicated to specific experimental models. Almost all research in molecular biology involves less than ten species, with the mouse *Mus musculus* and the zebrafish *Danio rerio* being the "universal" vertebrates (Dooley and Zon 2000; Sharpless and DePinho 2006), the cress *Arabidopsis thaliana* being the "universal" plant (Woodward and Bartel 2018), the bacterium *Escherichia coli* and the yeast *Saccharomyces cerevisiae* being the "universal" microorganisms (Orr-Weaver et al. 1981; Lee and Lee 2003), and the worm *Caenorhabditis elegan*s and the fly *Drosophila melanogaster* being the "universal" invertebrates (Rankin et al. 1990; Tickoo and Russell 2002). In contrast, there are thousands of species currently studied by ecologists and evolutionary biologists. There is no universal model. The variety of species involved ranges from viruses (e.g., Bull, 2006) to whales (e.g., Alexander et al., 2016).

Microcosms are also used as models in, say, ecology and evolutionary biology. These are often experimental investigation of a trait in the laboratory or in the field; there are many thousands of examples, some dating back to the beginnings of these disciplines (e.g., Beal 1885; Lutz 1915; Sturtevant 1915). Increasingly, some of the same microcosms used are those used in biochemistry, cell biology, and molecular biology (Jarosz and Dudley 2017; Kawecki et al. 2018). However, an evolutionary biologist studying cells would investigate the influence of, say, natural selection on the rate of cell division, whereas the cell

biologist would study the mechanics of cell division. This distinction is often viewed by biologists as being a distinction between a focus on "why?" and a focus on "how?"; see Mayr (1961), Tinbergen (1963), Ariew (2003), Laland et al. (2011), Bateson and Laland (2013), and Nesse (2013).

The beliefs accompanying models in ecology and evolutionary biology are usually that differences among species are "signal" not "noise" and that different species provide unique information. One reason for the focus on differences is that they are causally central in evolutionary processes such as natural selection (cf., Murray 1991). These beliefs sustain careers dedicated to different kinds of organisms (e.g., herpetologists, mammalogists, and ornithologists, many of whom study one or a few species, which are often chosen because they are *not* studied by others).

Ecologists and evolutionary biologists also traffic in more mathematical models, especially those that are viewed as having broad relevance, than do cell, molecular, and systems biologists. Many ecologists and evolutionary biologists view results derived from mathematical models as providing conceptual insights that are central to their work, even if it is experimental. Examples in ecology of such mathematically-derived model insights are (Levins 1968*a*; MacArthur 1970); examples in evolutionary biology include Fisher (1930), Wright (1931), and Haldane and Jayakar (1963).

What underlies these sub-disciplinary differences in regard to the influence of mathematical models? It is not as though cellular and molecular processes are inherently un-amenable to mathematical analysis. The focus on the experimental microcosms in cell, molecular, and systems biology is largely a result of the influence of Thomas Hunt Morgan and of Jacques Loeb. Each did early widely-influential 20th century work in embryology, genetics, and physiology and advocated for a biology in which the use of experimental models is paramount (Allen 1978; Pauly 1987; Brush 2002). They trained or mentored many biologists who gained substantial influence and who in turn trained many more biologists (so much so that many thousands of current biologists are their scientific "descendants"; the first author is one of them).

A focus on experimental microcosms licenses career and institutional investment in work to master the techniques needed to yield interpretable results. These influences act to diminish the use of mathematical models. The

notion that broad insights are to be gained via the investigation of "universal" experimental models is paramount.

Sub-disciplinary differences in modeling are increasingly influenced by institutional structure. Some important institutions for biological research and training in the United States have few if any faculty members doing research in ecology and evolutionary biology; these include Brandeis University, California Institute of Technology, Johns Hopkins University, Massachusetts Institute of Technology, Northwestern University, Rockefeller University, and the Salk Institute for Biological Sciences. At several, a "Department of Biology" became solely focused on cell and molecular biology by the elimination of other sub-disciplines. We know of no examples in which a "Department of Biology" became solely focused on ecology and evolution by the elimination of other sub-disciplines. At other institutions, sub-disciplines have not been eliminated but the "Department of Biology" has fissioned into the "Department of Cell and Molecular Biology" and the "Department of Ecology and Evolutionary biology" (or name variants thereof). Both kinds of transformation sustain if not strengthen sub-disciplinary differences in modeling.

2. Sub-disciplinary similarities in the way biologists view the process of model building

Despite differences in types of model typically used, sub-disciplines of biology are similar in two ways. The first is that most practitioners have an aversion to discussion of the nature and practice of modeling. This is especially true of normative guidelines for modeling. Such an activity is often deemed to be "unscientific" and a "waste of time", an underlying trope being that "facts and experiments mean something, philosophizing does not". This attitude may in part be due to a lack of exposure to such material during training in biology, which often has little or no exposure to mathematics and statistics (cf., Bialek and Botstein 2004; Marshall and Durán 2018). By contrast, even undergraduate training in physics includes topics such as relativity and quantum mechanics that necessitate some exposure to "philosophizing" about modeling and about what observations mean.

The second way in which the sub-disciplines are similar is that few researchers have called for increased attention to what modeling is and should be.

In cell, molecular, and systems biology, such calls include Gunawardena (2014), Torres and Santos (2015), Bartocci and Lió (2016), Sztul (2017), and Medina (2018). Most appeal for more use of computational and mathematical models. These calls appear to be motivated by the massive increase in the availability of data concerning the genomic, metabolomic, proteomic, and transcriptomic "levels" of the organism. For example, just *one* study (Telenti et al. 2016) reports on 2.8×10^{13} nucleotides sequenced from 10,545 humans and documents 150 million sequence variants. Another motivation is the belief that the new data make it possible to understand biological "complexity". These calls are not just for the use of computers to help collect and store data. They are calls that the increased use of computational and mathematical models is *required* to provide new kinds of answers (e.g., Berro 2018; Wallmeier 2018). Whether or not this is true, it is too early to tell whether these calls will be widely-heeded and to what extent this increased use provides answers that would be much more difficult or impossible to attain via the use of experimental models.

In ecology and evolutionary biology, calls for increased attention to what modeling is and should be have a longer history. Prominent among these are Holling (1964) and Levins (1966), who claimed that generality (termed "breadth" by Holling), precision, and realism are three desired attributes of a model in population biology (and of models in general). Both authors discussed the tradeoff among these attributes, with Holling claiming that modifying a model so as to increase any two, say, generality and realism, need not decrease precision, and Levins claiming that such a decrease is unavoidable. A necessary tradeoff implies that there are three types of models, which differ in which attribute is sacrificed in order to increase the other two. Levins viewed models with more generality and realism and less precision as most desirable. He did not define how generality, realism, and precision of a model can be assessed or demonstrate that they exhibit a necessary tradeoff. Levins further claimed that (p. 422) "Our truth is the intersection of independent lies", i.e., we can regard the common prediction arising from multiple "independent" models as "truth".

More than fifty years have passed since the publication of Levins' article and it has received over 1900 Google Scholar citations as of early 2019; there are far fewer citations of Holling's article, which was not cited by Levins. What can we conclude from this large number of citations of Levins' article? Does it mean that his claims have improved modeling in biology and that they should continue to provide guidance?

The answer to this question is unclear. Levins' article has often been cited in ways that suggest misunderstanding of or disagreement with his claims (see Orzack and Sober 1993, who reviewed all citations of his article up to 1993, and Orzack 2005). For example, Armstrong (1988) claimed that his model of an ecological community *is* general, realistic, and precise. Most often claims about model attributes are ambiguous because they are not anchored in explicit comparisons of models (as in "model x is more realistic than model y"). Many citing authors describe their model as being the type preferred by Levins, which at least naively can be taken to imply belief in the existence of his trichotomy of models, but what features of their model make it this type are not specified. The character of most citations of Levins' paper suggests (but does not prove) that they are mainly an effort to provide "quasi-philosophical" support for a model apart a demonstration that it provides biological insights.

Much of this history reflects the attitudes of biologists a generation or more ago. It is still conceivable that Levins' claims are relevant to modeling in the 21st century. Do biologists believe this to be true? One way to judge this is to assess the recent biological literature as recorded in the Google Scholar database. For example, it lists approximately 20,800 articles published in 2017 that contain "biology" and "modeling" in their abstract or title. Of these, 78 cite Levins' article. These citations occur almost always in the context of population and evolutionary biology, even though his claims apply to any natural science and his article was published in a general science journal. For example, as of March 2018, Google Scholar lists 366 post-2013 articles that have "biology" and "modeling" in the abstract and which cite Levins' article. The majority of these articles present original biological research; just two do not concern population or evolutionary biology (Shirsat et al. 2015; Ho et al. 2018). (Other citations are by philosophers discussing Levins' ideas or are by other kinds of scientists).

We do not know whether non-citation is mostly the result of scientists being unaware of Levins' article or that they deem it irrelevant or wrong. (It is even possible that non-citation occurs because it is regarded as self-evidently correct.) What is clear is that few biologists (much less other scientists) cite Levins' article and that almost all biologists that do cite it study population or evolutionary biology. Perhaps their non-citing colleagues are ignorant; perhaps they are enlightened.

The apparently very small constituency of Levins' article among biologists raises questions about its constituency among philosophers of biology and biologists concerned with the philosophical implications of modeling in biology. Critical assessment of his article began with Orzack and Sober (1993). Some commentators believe that Levins' article provides important insights and normative claims about modeling (Wimsatt 1987, 1981; Godfrey-Smith 2006; Plutynski 2006; Weisberg 2006*a*, *b*, 2007, 2013; Weisberg and Reisman 2008; Matthewson and Weisberg 2009; Goldsby 2013). Others are skeptical (Orzack and Sober 1993; Orzack 2005; Orzack 2012; Odenbaugh and Alexandrova 2011; Justus 2012). A main point of debate is Levins' claim about tradeoffs. Orzack and Sober (1993) showed that one can increase the generality and realism of a model without a decrease of its precision (see responses by Levins 1993 and Matthewson and Weisberg 2009).

This debate appears to have done little to help biologists develop better models, perhaps in part because the debate mainly involves philosophers of biology and journals typically not read by biologists (e.g., Biology & Philosophy). Almost all citations by biologists are at most acknowledgements that there is a debate (e.g., Martínez del Rio 2008) although some biologists engage with its substance (e.g., Slobodkin 1994). Perhaps some have been spurred to create models with increased generality, realism, and precision as compared to previous models after reading Orzack and Sober's demonstration that tradeoffs are not inevitable. If so, they have done this without citing Levins (1966) or Orzack and Sober (1993).

Many of the philosophers engaged in the debate over Levins' claims appear to have limited knowledge about the practice of modeling in biology. Instead, they appear to derive their knowledge about modeling from Levins' article. They also appear to be unaware of the very small constituency that Levins'

claims have among biologists and of how much biology has changed since the 1960s (see above), which reduces the practical relevance of Holling's and Levins' claims. Better knowledge of current biology would likely improve the philosophical and normative content of the debate.

We expect that the lack of connection between practitioners of biology and those who could provide useful conceptual and normative guidance will remain if not increase, given the ongoing avalanche of ecological and molecular data, which underlies claims that the 21st century is the century of biology (Venter and Cohen 2004; National Research Council 2009) and that there will be breakthroughs in, for example, the treatment of diseases and the remediation of environmental degradation. This possibility might cause biologists to be more open to normative guidelines derived from more formal consideration of what modeling is and should be. However, this appears not to be true as of yet.

The future usefulness to biologists of Holling's and Levins' claims is unclear. Perhaps they will provide normative guidance, even if this only amounts to better awareness of how model attributes such as generality, realism, and precision relate to one another. That said, their claims are rooted in the doing of the data-poor biology of fifty years ago. Their claims might be justified as an attempt to understand complex systems in the relative absence of data. Whatever relevance their insights still have, current biologists have abundant data and resources needed to analyze them such as computers, databases, and statistical and mathematical tools. Biologists and philosophers of biology hoping to provide much-needed normative guidance to biologists using models must pay attention to the realized character of current data and tools, not those of the 1960s.

One consequence of the availability of data is that the goal of biological modeling is increasingly the identification of a model that makes a *non-robust* prediction. In particular, biologists often seek a model is tailored to fit the specific biology under investigation and is not necessarily useful outside of this domain. Predictions are not derived from an ensemble of models (cf. Levins 1966, p. 423). There appears to be a diminishing potential constituency for Levins' claim that the identification of such robust model predictions is a good way to discover biological truth. This search for models that are non-robust has

been aided by the development of the Akaike Information Criterion, which can be used as a basis for choosing which model from a set of plausible models has the greatest support from the data (Akaike 1974; Burnham and Anderson 2002; Johnson and Omland 2004).

The search for a model with a non-robust prediction occurs in many sub-disciplines of biology. It has become especially common in ecology and evolutionary biology. For example, the determination of a phylogeny or evolutionary "tree" for a group of species given data on the trait expressed by each species now routinely involves finding the instantiated evolutionary model that makes the data most likely (e.g., Felsenstein 2004; Lemey et al. 2009). The investigator often does not even choose the uninstantiated model of the trait's evolution and so it is unclear in what sense (s)he is aware of potential tradeoffs among models.

As cell, molecular, and systems biology encompass more and more data, there has also been more explicit formulation of computational and mathematical models that make non-robust predictions. For example, Altan-Bonnet and Germain's (2005) mathematical model of signaling in the immune system (see their Figures 2B and S6) is based upon a "complete" representation of the signaling pathway underlying the T-cell antibody response. This representation, which includes hundreds of reaction steps, is based upon experimental work to elucidate the pathway. As such, the predictions of this model are non-robust in as much as it is intended to exactly represent the biology under investigation. There is no search for a prediction that is common to several models. As in the case of experimental models in these sub-disciplines of biology, the expectation is that the predictions are relevant to many organisms, not just those from which the experimental data were derived.

3. Causal models, data models, and algorithmic models

The search for non-robust predictions is part of a broader move by scientists to focus on algorithms and data as hypothesis "generators". This tendency has been addressed by Breiman (2001) who distinguished between the "data modeling culture" and the "algorithmic modeling culture". In the former, the

scientist chooses an underlying statistical model that (s)he believes generated the data and then uses it to make inferences about the observed data (e.g., whether the arithmetic average of an observed sample of data has a low or high probability of occurrence given random sampling of data generated by the chosen statistical model). In the latter, the scientist generates "black box" predictions from the data, without recourse to an initial choice of an underlying statistical model.

Breiman's distinction is couched in terms of statistics and so it does not describe what one could call the "causal modeling culture" in biology, in which idealized features of a biochemical pathway, or an organism, or a population, or a ecosystem, etc., are used to create the model. For example, when an ecologist creates a model of population dynamics, (s)he chooses whether or not there is a single population, whether or not a population is finite, whether or not two sexes are present, whether or not individuals mate randomly, whether or not the number of offspring produced is finite, whether or not there is environmental variation, and so forth (see example below). Statistical analysis of data does not inform the choice of model alternatives offered (although the choice of a particular alternative on offer is sometimes informed by statistical analysis). Hypothesis testing need not be the goal of the model. When done, it requires assumptions extrinsic to the assumptions used to create the model. In contrast, hypothesis testing of "statistical" and "algorithmic" models is intrinsic to the model (either because an assumption is made about the error distribution or the data are used to generate the distribution).

This distinction is underscored by the fact that some "causal" model assumptions could never be supported by data. For example, it is never true that, say, a population has infinite size, that individuals have an infinite number of offspring, that the environment is constant, etc. Nonetheless, one or more of these assumptions have long been used in the creation of "causal" models in ecology, many of which have yielded important insights.

In sub-disciplines dominated by experimental models, the "causal modeling culture" has led to some of the great discoveries of 20th century biology. For example, Meselson and Stahl (1958) created two distinct causal models in order to identify the correct mechanism for the replication of the DNA helix. However, data models play an increasing role in these sub-disciplines. For

example, mutations in a DNA sequence are often identified as being potentially disease-causing solely by the strength of their statistical association with the disease and not by their consequences for the function of the associated protein or of the associated biochemical pathway (Balding 2006; Edwards et al. 2013).

In sub-disciplines in which mathematical models are widely used, the use of "causal" models is common. For example, in evolutionary biology, there is a well-known model that is intended to explain the female-biased sex ratios found in some insect populations (Hamilton 1967). The model predicts that a female-biased sex ratio is the most evolutionarily-beneficial or "optimal" when a finite number of fathers and mothers comprise a "local" mating group. (Other assumptions are also required.) This prediction is based upon idealized biology that can never be true, e.g., that a mother produces an infinite number of offspring. (The prediction changes when offspring number is assumed to be finite, see Nagelkerke 1996). Hamilton's model is not a "data" model or an "algorithmic" model; the prediction of the model is not derived from statistical analysis of the relationship between the sex ratio and the number of fathers and mothers in a mating group.

Other important causal models in evolutionary biology have led the reconciliation of the genetic mechanisms underlying traits that vary discretely and those that vary continuously (Fisher 1918) and to the development of population-genetic models that include deterministic and stochastic evolutionary forces (e.g., Wright 1931). In turn, both of these led to the development of a highly influential but still controversial "synthetic" causal model that connects short-term and long-term evolution (e.g., see Laland et al. 2014; Wray et al. 2014).

The "causal modeling culture" and "data modeling cultures" are very different. Bringing them together is not just a matter of "adding" standard assumptions about sampling error to causal models (see below). Unfortunately, the important distinctions between the two cultures are often misunderstood. For example, Gunawardena (2014) confounds the two in his overview of modeling in biology when he writes (p. 6) that "Judging from some of the literature, we seem to forget that a model does not predict the data to which it is fitted: the model is chosen to fit them." Similarly, Nijhout et al. (2015) write (p. 2) that "… the 'model' is not a fixed object, but continually evolves through testing it

against data and revising it accordingly." In fact, causal models as defined here are *very rarely* formed and re-formed by data in the direct sense these authors describe (although observed data may inspire the creation of a causal model). Their descriptions concern the "data modeling culture" although the models they discuss are products of the "causal modeling culture".

Although the "causal modeling culture" and the "data modeling culture" have long histories in biology, they are problematic in important ways. An important problem for causal models can be illustrated in the context of the optimality model in evolutionary biology mentioned above. Consider a deviation between the optimal sex ratio predicted by Hamilton's model and an observed sex ratio. How large can such a deviation be and still allow us to conclude that the model provides a causal explanation of the data? The deviation cannot automatically be assumed to be due just to sampling error because it has causal implications. All other things being equal, the deviant observed sex ratio has a lower evolutionary fitness than does the optimal sex ratio and thereby should not be observed in the population. Is the observed sex ratio optimal but not correctly specified by an incorrectly-formulated model? Or is the model correctly formulated but the optimal sex ratio cannot evolve in the population studied? Neither can be assumed to be true a priori.

These central issues concerning how to understand the relationship between causal model predictions and data are almost never explored. The consequence has been inferential ambiguity because there are no agreed-upon standards by which success and failure are judged. The criteria used are often private and apparently arbitrary. For example, the *same* test of the predictions of an optimality model has engendered these opposite assessments: "there is a striking correspondence between theory and data" and "there is a great deal of scatter around the quantitative prediction" (see Orzack 2014 for details). These statements were based solely on visual inspection of the predictions and data, which is known to be strongly influenced by the graphical presentation (Cleveland and McGill 1987). Each of these assessments licensed opposite conclusions as to whether the observed data indicate that the species possesses an optimal trait. Sometimes a "qualitative" test of model predictions is used in which a predicted trend (e.g., downward) is compared with the observed trend of the data. This is not inherently problematic, although it leaves unresolved

what causal conclusions one should draw from concordance or discordance between the predicted and observed trends. These ambiguities in regard to the means by which the correspondence between data and causal model predictions is interpreted have received almost no attention by biologists; they are addressed in the context of testing optimality models by Orzack (1990), Orzack et al. (1991), and Orzack and Sober (1994).

The "data modeling culture" is also problematic but in part for different reasons. As noted by Breiman (2001), the assumptions routinely made about the model underlying the data are often of unknown validity at best. In addition, the criteria by which associations are judged to be "significant" or not are arbitrary and may often lead to incorrect conclusions (Ioannidis 2005).

We note that the distinction we make between "causal models" and "data models" is not meant to imply that the former are the only way to gain causal understanding. Shipley (1999) and Spirtes et al. (2000) present a method for making causal inferences from "data models". See also Pearl (2009). Their important motivation is that one often is confronted with the need to make causal inferences when controlled experiments are difficult or impossible. Their method is little-used in biology and wider implementation is required before we can assess how often it produces biologically-meaningful causal insights. If it often does so (despite important problems with their statistical approach, see Karlin et al. 1983; Freedman 1987, 1997), their method will be important, if only because it could partially reduce the need for experimental intervention to infer causation.

The "causal modeling" and "data modeling" cultures continue to play prominent and often useful roles in modeling in most if not all sub-disciplines of biology. That said, as Breiman (2001) noted, there is an increasing "algorithmic modeling culture". An important manifestation of this culture is the use of machine learning in the analysis of data. Here, automated procedures are used to make predictions from data. The perceived advantage of these methods is that they allow the investigator to forego much of the hard work needed to construct, analyze, and validate a causal model and or a statistical model. The use of machine learning is nicely illustrated by Olden et al. (2008) who used three different approaches (classification and regression trees, artificial neural networks, and evolutionary algorithms) to investigate the causes

of heterogeneity in the number of fish species found in over 8,000 freshwater lakes in Canada. The results of all approaches suggest that important determinants of the number of species are the amount of precipitation, the length of the shoreline, and the area of the lake. Their analyses exemplify the power of the "algorithmic modeling culture", especially that a huge amount of data have been encompassed in the analyses, the analyses are carried out with software that is readily available, a causal model has not been created "from scratch", and testable predictions about causes have been specified. As Olden et al. (2008) note, these methods have their weaknesses (such as the potential of over-fitting; see their Table 1). That said, these methods allow biologists to tackle analyses that were previously un-addressable. The importance of this cannot be exaggerated.

In this context, it is worth noting that Levins' claims about model building were in part intended to counter the view that ecologists need to focus on data analysis and numerical simulation so as to understand "systems ecology" (Levins 1968*b*; Palladino 1991). Levins claimed that this approach could not lead to causal understanding because (p. 421) there are "too many parameters to measure", the "equations are insoluble analytically", and any predictions would "have no meaning for us". However, none of these is necessarily true for ecological models, even those in systems ecology. Of course, there are models for which there are "too many parameters to measure". But for some, there are sufficient data to allow all parameters to be estimated. Of course, there are models for which "equations are insoluble analytically". But for some, the equations are soluble or can be solved via analytical approximation. Of course, there could be model predictions that "have no meaning for us". But for some, predictions have meaning (or are eventually understood). Despite all of these possible difficulties, there is no basis for a claim that a particular approach to modeling could not result in causal understanding. It is Levins' absolute claim that lacks substantiation. He correctly identified potential difficulties but that does not inform how often they occur. In addition, Levins failed to mention that these difficulties may arise in the approach to model building that he advocates.

It is also worth noting that our ability to solve some of these difficulties is not static. For example, machine learning analyses in many sub-disciplines

in biology demonstrate that prediction and causal understanding of many-dimensional "complex" systems can be attained even when many but not all parameters can be measured and when likely-relevant equations are insoluble (e.g., see Shan et al. 2006; Tarca et al. 2007; Wernick et al. 2010; Kampichler et al. 2010; Touw et al. 2012; Sommer and Gerlich 2013; Schrider and Kern 2015, 2016, 2018; Dumancas et al. 2017; Dietze et al. 2018; Flassig and Schenkendorf 2018; Ghosal et al. 2018).

Although Levins' claim is false as description of what is necessarily true about models, it does serve as a reminder to not view any approach to data analysis and causal understanding as "automatic". After all, different machine learning approaches can yield different results (e.g., Olden et al. 2008; Kampichler et al. 2010). Biological judgment will always be needed.

A development related to the "algorithmic modeling culture" is the computational "reverse engineering" of biological systems. Here, data and "candidate" symbolic representations of the dynamics are iteratively combined so as to ultimately generate "the" equations underlying the dynamical system (Bongard and Lipson 2007; Schmidt and Lipson 2009; Brunton et al. 2016). This approach to model generation underscores how far biological modeling can go beyond apparent limitations, such as those identified by Levins. It has remarkable potential, even it is not "automatic" and must be applied with judgment.

4. Whither Biology?

It is unclear as to whether there will be "movement" to close the gaps among sub-disciplines of biology in regard to the culture of modeling and also to close the gap between biologists and philosophers and others interested in biological modeling. In the former case, it is encouraging that sub-disciplines that have very different histories in regard to modeling at least have in common a search for models with non-robust predictions (although it is unlikely that the disparate groups of biologists understand that this is a common goal). In the latter case, if biologists and philosophers of biology are to provide normative guidance so as to improve the practice of biological modeling, they must

situate their guidance in the context of current practice of biology. The biology of today is not the biology of twenty years ago, much less fifty years ago.

We have described some of the extraordinary changes that all sub-disciplines in biology have undergone because of unprecedented increases in data and computational resources. As daunting as it can seem to biologists to assimilate new data and techniques, it is likely that the most important task faced by biologists is being open to change in regard to what biological "complexity" is conceived to be.

Complexity is a badge of honor for many biologists. In many sub-disciplines, biologists use it to represent the notion that systems under investigation, whether they be biochemical pathways, organs, or ecosystems, are tangled inscrutable webs of interactions. The trope is that we can only barely understand the simplest aspects of such systems and that we will *always* fail in regard in our attempts to provide complete understanding. This attitude commonly coexists with the notion that the only way to achieve partial understanding is a reductionist approach that involves "disassembling" the system. This approach is regarded as necessary but never sufficient to understand complexity.

In fact, this combination of attitudes is not inherently problematic and will continue to be fruitful. That said, we emphasize the need to be open to new understanding of what biological complexity actually is. It is telling that natural systems that could be construed as similarly highly complex and inscrutable are understood in different ways by different kinds of scientists. Phenomena such as cellular metabolism and energy flow in ecosystems are "complex" to biologists. Phenomena such as climate dynamics are "complex" to physicists, meteorologists, and geophysicists. Yet, biologists view the complexity as a manifestation of a unknowable multitude of interactions of roughly equivalent magnitude, whereas physical scientists view the complexity as a manifestation of a multitude of interactions but dominated by just a few (e.g., Ditlevsen and Johnsen 2010; Cimatoribus et al. 2012). In effect, physical scientists view the complexity as *simple*. This is sometimes justified by claims about separation of time scales, with some processes being "fast" enough that they only add noise to the "slow" low-dimensional drivers of the system (cf., Ditlevsen and Johnsen 2010, p. 2). Correct or not, the concordance perceived

by physical scientists between the dimensionality of the system and the dimensionality of the analysis is desirable at least in terms of inferential consistency. The discordance in this regard on the part of biologists is problematic and they would do well to eliminate it. In some instances, this might mean embracing complexity in the full sense (and abandoning low-dimensional tools) and in others by viewing complexity as simple and taking simple models seriously as providers of casual explanation (e.g., Reynolds 1987).

The contrast between biologists and physical scientists in attitudes towards the nature of complexity suggests that biologists may make substantial conceptual progress by being open to the possibility that complexity can be tractable. It is promising in this context to note that biologists studying "complex" systems they view as dynamically complex and physical scientists studying "complex" systems they view as dynamically simple often use the *same* tools from deterministic dynamical systems theory (e.g., see Strogatz 2018). The promise of this overlap is that it may help biologists change their understanding of how to model and understand biological complexity.

Acknowledgements

We thank an anonymous reviewer, Bruce Glymour, Elliott Sober, and Krzysztof Nowak-Posadzy for assistance and comments.

References

Akaike, H. 1974: A new look at the statistical model identification. IEEE Transactions on Automatic Control AC-19:716–723.

Alexander, A., D. Steel, K. Hoekzema, S. Mesnick, D. Engelhaupt, I. Kerr, R. Payne, and C.S. Baker. 2016: What influences the worldwide genetic structure of sperm whales (Physeter macrocephalus)? Molecular Ecology 25:2754–2772.

Allen, G.E. 1978: Thomas Hunt Morgan: The Man and His Science. Princeton University Press, Princeton.

Altan-Bonnet, G., and R.N. Germain. 2005: Modeling T cell antigen discrimination based on feedback control of digital ERK responses. PLoS Biology 3:e356.

Ariew, A. 2003: Ernst Mayr's "ultimate/proximate" distinction reconsidered and reconstructed. Biology & Philosophy 18:553–565.

Armstrong, R.A. 1988: The effects of disturbance patch size on species coexistence. Journal of Theoretical Biology 133:169–184.

Balding, D.J. 2006: A tutorial on statistical methods for population association studies. Nature Reviews Genetics 7:781–791.

Balme, D.M. 1987: The place of biology in Aristotle's philosophy. Pp. 9–20 *in* A. Gotthelf and J.G. Lennox, eds. Philosophical Issues in Aristotle's Biology. Cambridge University Press, Cambridge.

Barbour, I.G. 1974: Myths, Models, and Paradigms: A Comparative Study in Science and Religion. Harper & Row, New York.

Bartlett, M.S. 1956: Deterministic and stochastic models for recurrent epidemics. Pp. 81–109 *in* J. Neyman, ed. Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability Volume 2: Contributions to Probability Theory. Vol. 4. University of California Press, Berkeley.

Bartocci, E., and P. Lió. 2016: Computational modeling, formal analysis, and tools for systems biology. PLOS Computational Biology 12:e1004591.

Bateson, P., and K.N. Laland. 2013: Tinbergen's four questions: an appreciation and an update. Trends in Ecology & Evolution 28:712–718.

Beal, W.J. 1885: The viability of seeds. Proceedings of the Society for the Promotion of Agricultural Science 5:44–46.

Berro, J. 2018: "Essentially, all models are wrong, but some are useful"—a cross-disciplinary agenda for building useful models in cell biology and biophysics. Biophysical Reviews:10(6): 1637–1647.

Bialek, W., and D. Botstein. 2004: Introductory science and mathematics education for 21st-century biologists. Science 303:788–790.

Bongard, J., and H. Lipson. 2007: Automated reverse engineering of nonlinear dynamical systems. Proceedings of the National Academy of Sciences 104:9943–9948.

Breiman, L. 2001: Statistical modeling: the two cultures (with comments and a rejoinder by the author). Statistical Science 16:199–231.

Brunton, S.L., J.L. Proctor, and J.N. Kutz. 2016: Discovering governing equations from data by sparse identification of nonlinear dynamical systems. Proceedings of the National Academy of Sciences 113:3932–3937.

Brush, S.G. 2002: How theories became knowledge: Morgan's chromosome theory of heredity in America and Britain. Journal of the History of Biology 35:471–535.

Bull, J.J. 2006: Optimality models of phage life history and parallels in disease evolution. Journal of Theoretical Biology 241:928–938.

Burnham, K.P., and D.R. Anderson. 2002: Model Selection and Multi-model Inference: A Practical Information-theoretic Approach. Springer, New York.

Byrne, H.M. 2010: Dissecting cancer through mathematics: from the cell to the animal model. Nature Reviews Cancer 10:221–230.

Cimatoribus, A.A., S.S. Drijfhout, V.N. Livina, and G. Van der Schrier. 2012: Dansgaard-Oeschger events: tipping points in the climate system. Clim Past Discuss 8:4269–4294.

Cleveland, W.S., and R. McGill. 1987: Graphical perception: The visual decoding of quantitative information on graphical displays of data. Journal of the Royal Statistical Society. Series A (General) 150:192–229.

Colbert, E. 1962: The weights of dinosaurs. American Museum Novitates 2076:1–16.

Committee on a New Biology for the 21st Century. 2009: A New Biology for the 21st Century. The National Academies Press, Washington, DC.

Dietze, M.C., A. Fox, L.M. Beck-Johnson, J.L. Betancourt, M.B. Hooten, C.S. Jarnevich, T.H. Keitt, M.A. Kenney, C.M. Laney, L.G. Larsen, H.W. Loescher, C.K. Lunch, B.C. Pijanowski, J.T. Randerson, E.K. Read, A.T. Tredennick, R. Vargas, K.C. Weathers, and E.P. White. 2018: Iterative near-term ecological forecasting: Needs, opportunities, and challenges. Proceedings of the National Academy of Sciences 115:1424–1432.

Ditlevsen, P.D., and S.J. Johnsen. 2010: Tipping points: early warning and wishful thinking. Geophysical Research Letters 37:L19703.

Dooley, K., and L.I. Zon. 2000: Zebrafish: a model system for the study of human disease. Current Opinion in Genetics & Development 10:252–256.

Dumancas, G.G., I. Adrianto, G. Bello, and M. Dozmorov. 2017: Current developments in machine learning techniques in biological data mining. Bioinformatics and Biology Insights 11:1–4.

Edwards, S.L., J. Beesley, J.D. French, and A.M. Dunning. 2013: Beyond GWASs: illuminating the dark road from association to function. The American Journal of Human Genetics 93:779–797.

Egerton, F.N. 1975: Aristotle's population biology. Arethusa 8:307–330.

———. 2001: A history of the ecological sciences, part 2: Aristotle and Theophrastos. Bulletin of the Ecological Society of America 82:149–152.

Felsenstein, J. 2004: Inferring phylogenies. Sinauer Associates, Sunderland, MA.

Fisher, R.A. 1918: The correlation between relatives on the supposition of Mendelian inheritance. Transactions of the Royal Society of Edinburgh 52:399–433.

———. 1930: The Genetical Theory of Natural Selection. Clarendon Press, Oxford.

Flassig, R.J., and R. Schenkendorf. 2018: Model-based design of experiments: where to go? Ninth Vienna Internatioal Conference on Mathematical Modelling:875–876.

Freedman, D.A. 1987: As others see us: A case study in path analysis. Journal of Educational and Behavioral Statistics 12:101–128.

———. 1997: From association to causation via regression. Advances in Applied Mathematics 18:59–110.

Frigg, R., and S. Hartmann. 2018: *Models in Science*. Stanford Encyclopedia of Philosophy. Downloaded from https://plato.stanford.edu/archives/sum2018/entries/models-science/%7D Metaphysics Research Lab, Stanford University.

Ghosal, S., D. Blystone, A.K. Singh, B. Ganapathysubramanian, A. Singh, and S. Sarkar. 2018: An explainable deep machine vision framework for plant stress phenotyping. Proceedings of the National Academy of Sciences 115:4613–4618.

Godfrey-Smith, P. 2006: The strategy of model-based science. Biology & Philosophy 21:725–740.

Goldsby, M. 2013: The "structure" of the "strategy": Looking at the Matthewson-Weisberg trade-off and its justificatory role for the multiple-models approach. Philosophy of Science 80:862–873.

Gotthelf, A. 1999: Darwin on Aristotle. Journal of the History of Biology 32:3–30.

———. 2012: Teleology, First Principles, and Scientific Method in Aristotle's Biology. Oxford University Press, Oxford.

Gunawardena, J. 2014: Models in biology: "accurate descriptions of our pathetic thinking." BMC Biology 12:1–11.

Haldane, J.B. S., and S.D. Jayakar. 1963: Polymorphism due to selection of varying direction. Journal of Genetics 58:237–242.

Hamilton, W.D. 1967: Extraordinary sex ratios. Science 156:477–488.

Ho, P.-Y., J. Lin, and A. Amir. 2018: Modeling cell size regulation: From single-cell level statistics to molecular mechanisms and population level effects. Annual Review of Biophysics 47:251-271.

Hodgkin, A.L., and A.F. Huxley. 1952: A quantitative description of membrane current and its application to conduction and excitation in nerve. The Journal of Physiology 117:500–544.

Holling, C.S. 1964: The analysis of complex population processes. The Canadian Entomologist 96:335–347.

Ioannidis, J.P. A. 2005: Why most published research findings are false. PLOS Medicine 2:e124.

Jarosz, D.F., and A.M. Dudley. 2017: Meeting report on experimental approaches to evolution and ecology using Yeast and other model systems. G3: Genes|Genomes|Genetics 7:3237–3241.

Johnson, J.B., and K.S. Omland. 2004: Model selection in ecology and evolution. Trends in Ecology & Evolution 19:101–108.

Justus, J. 2012: The elusive basis of inferential robustness. Philosophy of Science 79:795–807.

Kacser, H., and J.A. Burns. 1973: The control of flux. Symposium of the Society of Experimental Biology 27:65–104.

Kampichler, C., R. Wieland, S. Calmé, H. Weissenberger, and S. Arriaga-Weiss. 2010: Classification in conservation biology: a comparison of five machine-learning methods. Ecological Informatics 5:441–450.

Karlin, S., E.C. Cameron, and R. Chakraborty. 1983: Path analysis in genetic epidemiology: a critique. American Journal of Human Genetics 35:695–732.

Kawecki, T.J., R.E. Lenski, D. Ebert, B. Hollis, I. Olivieri, and M.C. Whitlock. 2018: Experimental evolution. Trends in Ecology & Evolution 27:547–560.

Laland, K.N., K. Sterelny, J. Odling-Smee, W. Hoppitt, and T. Uller. 2011: Cause and effect in biology revisited: is Mayr's proximate-ultimate dichotomy still useful? Science 334:1512–1516.

Laland, K.N., T. Uller, M.W. Feldman, K. Sterelny, G.B. Müller, A.P. Moczek, E. Jablonka, and J. Odling-Smee. 2014: Does evolutionary theory need a rethink? Yes, urgently. Nature 514:161–164.

Lee, P.S., and K.H. Lee. 2003: Escherichia coli—a model system that benefits from and contributes to the evolution of proteomics. Biotechnology and Bioengineering 84:801–814.

Lemey, P., M. Salemi, and A.-M. Vandamme. 2009: The Phylogenetic Handbook: A Practical Approach to Phylogenetic Analysis and Hypothesis Testing. Cambridge University Press, Cambridge.

Leroi, A.M. 2014: The Lagoon: How Aristotle Invented Science. Viking Penguin, New York.

Levins, R. 1966: The strategy of model building in population biology. American Scientist 54:421–431.

———. 1968a: Evolution in Changing Environments: Some Theoretical Explorations. Princeton University Press, Princeton.

———. 1968b: Ecological engineering: theory and technology. The Quarterly Review of Biology 43:301–305.

———. 1993: A response to Orzack and Sober: Formal analysis and the fluidity of science. Quarterly Review of Biology 68:547–555.

Lewontin, R.C. 1968: Biological models. Pp. 342–246 *in* P.P. Wiener, ed. Dictionary of the History of Ideas Volume 1. Scribner's, New York.

Lutz, F.E. 1915: Experiments with Drosophila ampelophila concerning natural selection. Bulletin of the American Museum of Natural History 34:605–624.

MacArthur, R.H. 1970: Species packing and competitive equilibrium for many species. Theoretical Population Biology 1:1–11.

Marshall, J.A., and P. Durán. 2018: Are biologists getting the mathematical training they need in college? Biochemistry and Molecular Biology Education 46:612–618.

Martínez del Rio, C. 2008: Metabolic theory or metabolic models? Trends in Ecology & Evolution 23:256–260.

Matthewson, J., and M. Weisberg. 2009: The structure of tradeoffs in model building. Synthese 170:169–190.

Mayr, E. 1961: Cause and effect in biology. Science 134:1501–1506.

Medina, M.Á. 2018: Mathematical modeling of cancer metabolism. Critical Reviews in Oncology / Hematology 124:37–40.

Meselson, M., and F.W. Stahl. 1958: The replication of DNA in Escherichia coli. Proceedings of the National Academy of Sciences 44:671–682.

Murray, B.G. J. 1991: Sir Isaac Newton and the evolution of clutch size in birds: a defense of hypothetico-deductive method in ecology and evolutionary biology. Pp. 343–180 *in* J.L. Casti and A. Karlqvist, eds. Beyond Belief: Randomness, Prediction and Explanation in Science. CRC Press, Boca Raton, Florida.

Nagelkerke, C.J. 1996: Discrete clutch sizes, local mate competition, and the evolution of precise sex allocation. Theoretical Population Biology 49:314–343.

Nesse, R.M. 2013: Tinbergen's four questions, organized: a response to Bateson and Laland. Trends in Ecology & Evolution 28:681–682.

Nijhout, H.F., J.A. Best, and M.C. Reed. 2015: Using mathematical models to understand metabolism, genes, and disease. BMC Biology 13:1–10.

Odenbaugh, J., and A. Alexandrova. 2011: Buyer beware: robustness analyses in economics and biology. Biology & Philosophy 26:757–771.

Olden, J.D., J.J. Lawler, and N.L. Poff. 2008: Machine learning methods without tears: a primer for ecologists. The Quarterly Review of Biology 83:171–193.

Orr-Weaver, T.L., J.W. Szostak, and R.J. Rothstein. 1981: Yeast transformation: a model system for the study of recombination. Proceedings of the National Academy of Sciences 78:6354–6358.

Orzack, S.H. 1990: The comparative biology of second sex ratio evolution within a natural population of a parasitic wasp, Nasonia vitripennis. Genetics 124:385–396.

———. 2005: What, if anything, is "The strategy of model building in population biology?" - A comment on Levins (1966) and Odenbaugh (2003). Philosophy of Science 72:479–485.

———. 2012: The philosophy of modelling or does the philosophy of biology have any use? Philosophical transactions of the Royal Society of London. Series B 367:170–180.

———. 2014: A commentary on "The Formal Darwinism Project": there is no grandeur in this view of life. Biology & Philosophy 29:259–270.

Orzack, S.H., and E. Sober. 1993: A critical assessment at Levins's The strategy of model building in population biology (1966). Quarterly Review of Biology 68:533–546.

———. 1994: Optimality models and the test of adaptationism. The American Naturalist 143:361–380.

Orzack, S.H., E.D. Parker, Jr, and J. Gladstone. 1991: The comparative biology of genetic variation of conditional sex ratio behavior in a parasitic wasp, Nasonia vitripennis. Genetics 127:583–599.

Palladino, P. 1991: Defining ecology: ecological theories, mathematical models, and applied biology in the 1960s and 1970s. Journal of the History of Biology 24:223–243.

Pauly, P.J. 1987: Controlling Life: Jacques Loeb & The Engineering Ideal in Biology. Oxford University Press, New York.

Pearl, J. 2009: Causality. Cambridge University Press, Cambridge.

Plutynski, A. 2006: Strategies of model building in population genetics. Philosophy of Science 73:755–764.

Rakison, D.H. 2018: Do 5-month-old infants possess an evolved detection mechanism for snakes, sharks, and rodents? Journal of Cognition and Development 19:456–476.

Rankin, C.H., C.D.O. Beck, and C.M. Chiba. 1990: Caenorhabditis elegans: a new model system for the study of learning and memory. Behavioural Brain Research 37:89–92.

Reynolds, C.W. 1987: Flocks, herds and schools: a distributed behavioral model. Computer Graphics 21:25–34.

Schmidt, M., and H. Lipson. 2009: Distilling free-form natural laws from experimental data. Science 324:81–85.

Schrider, D.R., and A.D. Kern. 2015: Inferring selective constraint from population genomic data suggests recent regulatory turnover in the human brain. Genome Biology and Evolution 7:3511–3528.

———. 2016: S/HIC: robust identification of soft and hard sweeps using machine learning. PLOS Genetics 12:e1005928.

———. 2018: Supervised machine learning for population genetics: a new paradigm. Trends in Genetics 34:301–312.

Shan, Y., D. Paull, and R.I. McKay. 2006: Machine learning of poorly predictable ecological data. Ecological Modelling 195:129–138.

Sharpless, N.E., and R.A. DePinho. 2006: Model organisms: The mighty mouse: genetically engineered mouse models in cancer drug development. Nature Reviews Drug Discovery 5:741–754.

Shipley, B. 1999: Testing causal explanations in organismal biology: causation, correlation and structural equation modelling. Oikos 86:374–382.

Shirsat, N.P., N.J. English, B. Glennon, and M. Al-Rubeai. 2015: Modelling of mammalian cell cultures. Pp. 259–326 *in* M. Al-Rubeai, ed. Animal Cell Culture. Springer, Cham, Switzerland.

Slobodkin, L.B. 1994: The connection between single species and ecosystems. Pp. 35–87 *in* D.W. Sutcliffe, ed. Water Quality and Stress Indicators in Marine and Freshwater Ecosystems: Linking Levels of Organisation (Individuals, Populations, Communities). Freshwater Biological Association, Ambleside.

Sommer, C., and D.W. Gerlich. 2013: Machine learning in cell biology–teaching computers to recognize phenotypes. Journal of Cell Science 126:5529–5539.

Spirtes, P., C. Glymour, and R. Scheines. 2000: Causation, Prediction, and Search. MIT Press, Cambridge.

Strogatz, S.H. 2018: Nonlinear Dynamics and Chaos: with Applications to Physics, Biology, Chemistry, and Engineering. CRC Press, Boca Raton, Florida.

Sturtevant, A.H. 1915: Experiments on sex recognition and the problem of sexual selection in Drosoophilia. Journal of Animal Behavior 5:351–366.

Szostak, J.W., T.L. Orr-Weaver, R.J. Rothstein, and F.W. Stahl. 1983: The double-strand-break repair model for recombination. Cell 33:25–35.

Sztul, E. 2017: How can biological modeling help cell biology? Cellular Logistics 4:e1404780.

Tarca, A.L., V.J. Carey, X. Chen, R. Romero, and S. Drăghici. 2007: Machine Learning and Its Applications to Biology. PLOS Computational Biology 3:e116.

Tcherkez, G.G. B., C. Bathellier, H. Stuart-Williams, S. Whitney, E. Gout, R. Bligny, M. Badger, and G.D. Farquhar. 2013: D2O solvent isotope effects suggest uniform energy barriers in Ribulose-1,5-bisphosphate Carboxylase/Oxygenase catalysis. Biochemistry 52:869–877.

Telenti, A., L.C. T. Pierce, W.H. Biggs, J. di Iulio, E.H. M. Wong, M.M. Fabani, E.F. Kirkness, A. Moustafa, N. Shah, C. Xie, S.C. Brewerton, N. Bulsara, C. Garner, G. Metzker, E. Sandoval, B.A. Perkins, F.J. Och, Y. Turpaz, and J.C. Venter. 2016: Deep sequencing of 10,000 human genomes. Proceedings of the National Academy of Sciences 113:11901–11906.

Tickoo, S., and S. Russell. 2002: Drosophila melanogaster as a model system for drug discovery and pathway screening. Current Opinion in Pharmacology 2:555–560.

Tinbergen, N. 1963: On aims and methods of ethology. Zeitschrift für Tierpsychologie 20:410–433.

Tipton, J.A. 2008: Aristotle's observations of the foraging interactions of the red mullet (Mullidae: Mullus spp) and sea bream (Sparidae: Diplodus spp). Archives of Natural History 35:164–171.

Torres, N. V, and G. Santos. 2015: The (mathematical) modelling process in biosciences. Frontiers in Genetics 6:1–9.

Torres, N. V, F. Mateo, E. Melendez-Hevia, and H. Kacser. 1986: Kinetics of metabolic pathways. A system in vitro to study the control of flux. Biochemical Journal 234:169–174.

Touw, W.G., J.R. Bayjanov, L. Overmars, L. Backus, J. Boekhorst, M. Wels, and S.A. F.T. van Hijum. 2012: Data mining in the Life Sciences with Random Forest: a walk in the park or lost in the jungle? Briefings in Bioinformatics 14:315–326.

Venter, J.C., and D. Cohen. 2004: The Century of Biology. New Perspectives Quarterly 21:73–77.

Wallmeier, H. 2018: Quantitative biology-a perspective for the life sciences' way into the future. Journal of Business Chemistry 15:78–92.

Weisberg, M. 2006a: Forty years of 'The strategy': Levins on model building and idealization. Biology & Philosophy 21:623–645.

———. 2006b: Robustness analysis. Philosophy of Science 73:730–742.

———. 2007: Three kinds of idealization. The Journal of Philosophy 104:639–659.

———. 2013: Simulation and Similarity: Using Models to Understand the World. Oxford University Press, Oxford.

Weisberg, M., and K. Reisman. 2008: The robust Volterra principle. Philosophy of Science 75:106–131.

Wernick, M.N., Y. Yang, J.G. Brankov, G. Yourganov, and S.C. Strother. 2010: Machine learning in medical imaging. IEEE Signal Processing Magazine 27:25–38.

Wimsatt, W.C. 1981: Robustness, reliability and overdetermination. Pp. 324–163. *in* R. Brewer and B. Collins, eds. Scientific Inquiry and the Social Sciences. Jossey-Bass, San Francisco.

———. 1987: False models as means to truer theories. Pp. 33–55 *in* M.N. Nitecki and A. Hoffman, eds. Neutral Models in Biology. Oxford University Press, Oxford.

Woodward, A.W., and B. Bartel. 2018: Biology in bloom: A primer on the Arabidopsis thaliana model system. Genetics 208:1337–1349.

Wray, G.A., H.E. Hoekstra, D.J. Futuyma, R.E. Lenski, T.F.C. Mackay, D. Schluter, and J.E. Strassmann. 2014: Does evolutionary theory need a rethink? No, all is well. Nature 514:161–164..
Wright, S. 1931: Evolution in Mendelian populations. Genetics 16:97–159.

Steven Hecht Orzack
Fresh Pond Research Institute
Cambridge, MA 02140
orzack@freshpond.org

Brian McLoone
National Research University
Higher School of Economics
Moscow, Russia
and
Fresh Pond Research Institute
Cambridge, MA 02140
brianbmcloone@gmail.com