The philosophy of modelling or does the philosophy of biology have any use?

Steven Hecht Orzack*

Fresh Pond Research Institute, Cambridge, MA 02140, USA

Biologists in search of answers to real-world issues such as the ecological consequences of global warming, the design of species’ conservation plans, understanding landscape dynamics and understanding gene expression make decisions constantly that are based on a ‘philosophical’ stance as to how to create and test explanations of an observed phenomenon. For better or for worse, some kind of philosophy is an integral part of the doing of biology. Given this, it is more important than ever to undertake a practical assessment of what philosophy does mean and should mean to biologists. Here, I address three questions: should biologists pay any attention to ‘philosophy’; should biologists pay any attention to ‘philosophy of biology’; and should biologists pay any attention to the philosophy of biology literature on modelling? I describe why the last question is easily answered affirmatively, with the proviso that the practical benefits to be gained by biologists from this literature will be directly proportional to the extent to which biologists understand ‘philosophy’ to be a part of biology, not apart from biology.

Keywords: ecology; philosophy; modelling; systems biology

1. INTRODUCTION

To many scientists, the phrase ‘the philosophy of modelling’ generates a feeling akin to that experienced when anticipating a visit to the dentist. At best, it is boring. More likely, it is painful. At least for many empirically focused scientists, ‘philosophy’ is rarely acknowledged, and when it is, it probably brings to mind a dreary, overly formal undergraduate course with topics such as metaphysics, scientific ‘logic’ or ontology. The summary attitude taken about the material is ‘perhaps worthwhile to those comfortably situated in armchairs, but of little or no use to practicing scientists’. Even some philosophers have this attitude. It is David Hume (see [1, p. 165]) who wrote ‘If we take in our hand any volume; of divinity or school metaphysics, for instance; let us ask, Does it contain any abstract reasoning concerning quantity or number? No. Does it contain any experimental reasoning concerning matter of fact and existence? No. Commit it then to the flames; for it can contain nothing but sophistry and illusion.’ This passage licenses a variety of interpretations, but here I take it to exemplify the kind of impatience many scientists have with ‘philosophical’ considerations as to the way to do science, especially that which might distract from the effort to discover and explain hard facts about nature. It is worth emphasizing in this context that almost all scientists are scientific ‘realists’; they believe that they are completely objective or nearly so and that their facts are real. Given this, time spent on anything other than the work of collecting facts and explaining them is time wasted.

But this is too glib an answer to the question as to the use of philosophy to scientists (independent of whether one agrees or disagrees with the common position about scientific realism). We need something more substantive. Scientists in search of answers to real-world issues such as the ecological consequences of global warming, the design of species’ conservation plans, understanding landscape dynamics and understanding gene expression [2–5] make decisions constantly that are based on a ‘philosophical’ stance as to how to do science. For better or for worse, some kind of philosophy is an integral part of the doing of science. Given this, it is more important than ever to undertake a practical assessment of what philosophy does mean and should mean to scientists. I now restrict myself to biologists and address three questions that concern if, why, and how philosophy has any use to biologists attempting to describe and predict biological systems, especially those with multiple causal levels.

The questions are: should biologists pay any attention to ‘philosophy’; should biologists pay any attention to ‘philosophy of biology’; and should biologists pay any attention to the philosophy of biology literature on modelling?

2. SHOULD BIOLOGISTS PAY ANY ATTENTION TO ‘PHILOSOPHY’?

This question is probably best parsed as relating to ‘philosophy’ as embodied in the content of introductory courses offered by the department of philosophy of a typical university or college. Most of the big topics one encounters in such a course have no direct application...
to the practice of science. On the other hand, one can readily see at least the potential for an introduction to formal logic in order to augment one’s capability of ‘logical’ thinking about how to structure experiments and analyses and about what inferences are permitted given a particular outcome. Perhaps, such an introduction was the ultimate genesis of, say, positive and negative controls as used in molecular biology or of Koch’s postulates [6]. Nonetheless, philosophers have no lock on the ability to think logically and the creation of ‘logical’ guidelines for the construction of models is well within the reach of biologists. Perhaps, all that is needed is some ‘homegrown’ philosophy (‘logic of modelling’ by biologists for biologists). At the normative level, I would conclude that biologists need not pay attention to philosophy in the narrow sense for special insights and tips as to how to do better science, at least if they are willing to do some hard thinking for themselves. But, this is a big ‘if’. Thinking critically is always hard and if developing the capacity to do so is made any easier at least initially by an introductory course on logic, so much the better. Why not use all the help one can get?

3. SHOULD BIOLOGISTS PAY ANY ATTENTION TO ‘PHILOSOPHY OF BIOLOGY’?

In order to answer this question, it is useful to distinguish between work by philosophers about biology and philosophical work by biologists. Some biologists have paid attention to some of the work by philosophers of biology (as published in journals such as Biology & Philosophy). However, many citations of this literature by biologists appear merely to provide cover for a pre-existing attitude and perhaps not to reflect a grasp of the substantive philosophical issues at stake. For example, Evans et al. [4] endorsed a well-known claim about model building in biology (see §4); they then acknowledge some critiques of it, but proceed to ignore them without explanation. Their decision to do so may ultimately turn out to be correct, but, at minimum, as a community of scholars, whether biologist or philosopher, we owe each other more explicit exposition of reasons for disagreement if we are to make joint progress at explaining the world. For most subdisciplines within biology, there is not even this amount of contact. Of course, there are some subdisciplines for which claims made by philosophers have influenced practice. Even in these subdisciplines (such as evolution-ary biology), the reception of relevant and important normative claims by philosophers [7,8] is mixed, with many practitioners paying no attention.

The reception of philosophical work produced by biologists is somewhat different. While still disparaged by some biologists (one occasionally hears comments about a ‘philosopause’), there are some philosophical contributions by biologists that have received substantial attention [9]. A few are taken to provide normative guidance. Of these, the 1966 article ‘The strategy of model building in population biology’ by Richard Levins [10] is viewed by many as providing special insight into how to create models. Levins’s article contains two central claims. The first is that unavoidable trade-offs between model attributes (generality, realism and precision) make it impossible to create a model that is maximally general, realistic and precise. As a consequence, there are only three types of models, which are defined by the three possible pairs of model attributes that are maximized. According to Levins, the preferable type of model is ‘type III’, one that emphasizes generality and realism, but is not as precise as it could be; for example, Levins claimed that his models of evolution in variable environments were of this desired type. Levins’s second central claim is that a prediction shared by many ‘independent’ models is ‘robust’ and thereby a truth about nature or at least more likely to be true.

Levins’s paper has been cited more than 490 times (Web of Science, October 2011); many of these citations include an assignment of a particular model to one of the three types described by Levins. Most are assigned to type III. Orzack & Sober [11] provided the first biological and philosophical critique of Levins’s claims. Their summary of his claims is as follows (p. 533):

Every scientific discipline confronts the problem of coping with nature’s complexity. If every scientific theory is selective in the details it chooses to characterize and if each introduces simplifying assumptions, it is only reasonable to wonder how theories can ever hope to describe nature as it really is …… [Levins’s] solution to this problem consists of two important claims. The first is that model building involves a necessary trade-off among generality, realism and precision. The second important claim involves the concept of robustness. Levins asserts that truths about nature can be revealed by finding ‘robust theorems’. He uses this term to refer to a proposition that is a joint consequence of independent models of the same biological phenomenon. Finding such theorems supplies an access to truths about nature that supplements the more familiar procedure of testing theoretical predictions with data.

This article, along with an invited response by Levins [12] and later comments and elaborations [13–24], constitutes one of the few sustained dialogues between biologists and philosophers of biology (see also [25]). This and the fact that it concerns the conceptual and practical issues that arise when creating scientific models make it an ideal-specific context in which to answer the final question posed above.

4. SHOULD BIOLOGISTS PAY ANY ATTENTION TO THE PHILOSOPHY OF BIOLOGY LITERATURE ON MODELLING?

Levins deserves credit for taking on some thorny problems. After all, given the potential complexity of biological phenomena, how does one decide in a principled manner what to account for and what to ignore when creating a model? In addition, is a prediction made by more than one model a special aid in the discovery of biological truths? Many of the answers formulated to both of these questions appear to be ad hoc and it is tempting to view them as being guided by tradition, rather than by scientific considerations. To this extent, Levins’s paper and the papers it has stimulated address important problems that deserve to be
considered seriously. This amounts to nothing less than an unqualified ‘Yes!’ in response to the question at hand. (Of course, other contributions to the philosophy of biology literature also motivate this affirmative response, e.g. Rykiel & Grant [26] which introduces an issue of Ecological Modelling devoted to ‘An evaluation of the role of theoretical models in ecology’, and Colyvan et al. [27]. Biology is also not alone among the natural sciences in regard to having stimulated a philosophical literature that is at least partially useful e.g. see Oreskes et al. [28] and Myung et al. [29], which introduces an issue of the Journal of Mathematical Psychology devoted to model selection.)

But saying yes is something different than endorsing Levins’s claims. A central question remains: did Levins get it right? In fact, despite the importance of his attempt to grapple with an immensely important issue, I believe that his two central claims are incorrect. In particular, there is no necessary trade-off between model attributes such that only three types of model are possible. In addition, there is no reason to have higher confidence in the truth of a shared or ‘robust’ prediction when compared with one that is not so. Given this, why bother with Levins [10] and related literature? In fact, the debate is important independently of whether Levins is right or wrong because at minimum it illustrates that some central everyday concepts relied upon by biologists when developing models are much more ambiguous than many biologists appear to understand. These ambiguities detract from our efforts to model natural systems and the first step towards eliminating them is recognizing them.

My statement that Levins’s two claims are incorrect is best understood by first defining some terms. Levins did not define the model attributes of generality, and precision, and reality in his paper. Consider the following definitions presented by Orzack & Sober [11, p. 534]:

— If one model applies to more real-world systems than another, it is more general.

— If one model takes account of more independent variables known to have an effect than another model, it is more realistic.

— If a model generates point predictions for output parameters, it is precise.

Given these definitions, Orzack & Sober [11, p. 536] noted,

Consider Levins’s claim that one can choose to construct a general and unrealistic model or an ungeneral and realistic model. Let’s assess this claim by considering two familiar models of the instantaneous rate of change of population size ($N$):

$$\frac{dN}{dt} = rN \quad (4.1)$$

and

$$\frac{dN}{dt} = rN + aN^2. \quad (4.2)$$

Here $r$ is the growth rate of the population (assumed to be constant), and $a$ is a constant. Model (4.1) is the so-called ‘density-independent’ model, while model (4.2) is often called the ‘density-dependent’ model. However, this label is not accurate in the present context. Model (4.2) is a uninstantiated model; it allows for density independence ($a = 0$) and density dependence ($a < 0$). Model (4.1) is a special case of model (4.2) in that any population described by the uninstantiated model (4.1) is also described by the uninstantiated model (4.2). So model (4.2) is more general than model (4.1). It is also true that every variable that potentially plays a causal role in model (4.1) also is a variable in model (4.2), but not conversely. So model (4.2) is more realistic than model (4.1). In this case, the two properties are necessarily associated; generality and realism are not model attributes that may be altered independently.

This means that it is not possible to, say, increase the generality of a uninstantiated model by making it less realistic. Increasing the generality increases the realism. Furthermore, Orzack & Sober [11, p. 536] went on to show that generality and realism are not necessarily associated when models are instantiated (i.e. when, say, the relationship between two components of a model is given a specific numerical form, as opposed to just being an abstract function). The minimal point is that Levins’s claims as to the relationship among model attributes requires a specification as to whether the models are instantiated or not. Without such a specification, Levins’s specific claim is ambiguous. In addition, when models are taken to be uninstantiated, at least, the trichotomy described by Levins cannot exist (see also Orzack [14] who illustrates how increasing the generality of an uninstantiated model increases its realism and its precision).

Levins’s second claim about ‘robustness’ lacks coherence when said to apply across models inasmuch as models of the same phenomenon are usually not logically independent (the truth of one model implies something about the truth of another) and the notion of statistical independence of models lacks definition (see [11, pp. 538–540]). Robustness is ultimately most generously understood within a model, i.e. as a description of how much a given model prediction changes depending on changes in the underlying input parameters. Even given this more narrow interpretation (one familiar to most modelers), there is no particular reason to assume that such an ‘invariant’ prediction is more likely true as a prediction about nature than one that is not. In addition, it is credible that the goal of modelling should be the production of non-robust predictions inasmuch as they allow biologists to more readily determine which model is correct; robust predictions make it harder to choose between models. Despite this, allusions by biologists and others to ‘robust’ predictions are common and such predictions are often viewed as desirable (cf. [17,30]). I suspect that very little of this practice is grounded in the experience of finding that robust conclusions are more often correct when predictions and data are compared. Instead, I believe it at least partially reflects the higher status some in the scientific community attach to the activity of synthesis when compared with what they deem to be more utilitarian endeavours (cf. [31]).

It is worth noting here that many biologists and other scientists appear to act inconsistently in their attachment of value to robustness. On the one hand, it is
deemed to be a virtue in many scientific contexts. On the other hand, it is telling to consider how readily most scientists reject some robust predictions. After all, consider the ‘robust’ prediction that there is some supernatural being or God in the universe. The vast majority of the more than six billion humans alive today believe in some form of supernatural being. In a general sense, one can construe this to be a ‘robust’ prediction in Levins’s original sense; it is a inference made in common by billions of different ‘independent models’ (each human’s belief system or ‘model’ used to ‘predict’ that God exists has some unique features).

Yet, this robust prediction is reflexively (and correctly) rejected by most scientists. At least some of them reject this prediction because of the attitude that ultimately only data count when it comes to model assessment. Consistent reliance on the latter attitude would be a tremendous improvement in the practice of biology.

Orzack & Sober [11] go on to illustrate how arbitrary the actual practice by biologists of assigning models to the three types defined by Levins has been. Typically, model assignments are not anchored in any explicit description of the reasons for the assignment or in an explicit description of the models being compared (see §5). This further limits the practical insights that Levins’s framework provides.

5. WHERE ARE WE?
This debate in the philosophy of biology literature has at least three substantial implications for modelling practice. First, descriptions of this or that biological model as being ‘general’ (or not) are extremely common; for example, Evans et al. [4] defied what they view as a preference for simple models, which are construed as being more general. As described in §4, generality and realism are comparative notions; absolute (non-comparative) definitions of these attributes are very difficult if not impossible to conceive of. But, virtually all assessments about these model attributes lack explicit reference to the set of models being compared. Consider how often in a seminar you have heard a statement about model generality (e.g. the Hardy–Weinberg model in population genetics or the logistic model of population growth is a ‘general’ model) and have it go without comment. Typically, one accepts such statements without explicitly having in mind another model for comparison, much less access to the comparison any one of your colleagues is making in his or her head. The point here is that claims for or against any given model are meaningful only when they are explicitly comparative. The next time you hear such a claim, look for the comparison.

The second implication stems from the fact that the debate over the validity of Levins’s claims and our claims continues; this is a live issue. There is use to this debate, and which way we finally come to understand whether and how trade-offs structure biological models and what robustness means have enormous implications for the process of modelling. Biologists should pay attention. For example, Levins [12] disagreed with our definitions of model attributes. For example, he notes that realism can be increased not just by taking ‘account of more independent variables’ (see §4) but by (p. 548) adding ‘new variables that mutually affect each other’ (and so are not independent) and by adding ‘a link between variables already present’ (thereby removing their independence). Although I view Levins’s statements as being based on an ungenerously narrow reading of the phrase ‘independent variable’, his clarifications appropriately illustrate the variety of choices available to modellers in order to increase the generality of a model when compared with that of another model.

More recently, Matthewson & Weisberg [18] defended the claim that a trade-off between precision and generality exists; their analysis is predicated upon conceiving of model attributes such as generality as having a quantifiable magnitude. It further depends on distinguishing between the generality of a given model and the generality of a set of models (generality is defined with respect to the number of systems to which the model applies, see §4). In addition, they describe two different kinds of generality, one defined with respect to the number of biological systems the model actually applies, and the other defined with respect to the number of systems to which the model potentially applies. They show that no trade-off between precision and both kinds of generality exists when a single model is considered. By contrast, trade-offs exist when one considers sets of models. There is some irony in contrasting their analysis, which seeks to provide a formal mathematical underpinning to Levins’s 1966 claim about trade-offs, with Levins’s [12, p. 547] disparaging comment that this kind of mathematical and logical ‘programme’ to achieve logical understanding of science via the discovery of ‘clear definitions, unambiguous categories, sharp measurement and the discovery of algorithms that could substitute for human judgement…. has been a failure, as indeed it had to be’.

Matthewson and Weisberg’s analyses [18] are correct given their assumptions. To this extent, it is essential to note that they confirm the claim of Orzack & Sober [11] that there is no necessary trade-off between precision and generality, inasmuch as this claim is understood to apply to an individual model. This is the assumed modelling context in Levins [10], and it is one that I believe is assumed by most biologists. There are two additional comments to make about Matthewson and Weisberg’s argument. The first is that the conception of generality defined with respect to sets of models may have normative use, but this remains to be determined. The second comment is that their argument is based on some constructs that minimally lack operational definition. After all, how does one meaningfully quantify the number of systems to which a model or set of models actually applies as opposed to potentially applies? Either of these qualitative measures seems very elusive, much less both. To this extent, whether their formal analysis provides accessible normative guidance to scientists also remains to be determined; I am sceptical that it will provide such guidance. I end by noting that the three-way interaction among generality, reality and precision proposed by Levins still lacks a coherent exposition, even one that is formal, much less practical.
Weisberg [17] has also defended the claim that robust predictions are often preferable, even if they are not necessarily more likely to be true. Inasmuch as this claim distinguishes between the robustness of a prediction and its truth, it codifies an acceptable, but nonetheless still arbitrary tool by which one might gain biological understanding. The notion of models being ‘independent’ of one another still lacks coherence, but one might generously interpret Levins’s original usage of this word as referring to a joint prediction arising from a set of models, each of which has one or more features that it does not share with the rest of the set. Seen in this light, Levins’s1966 claim about models being ‘independent lies’ lacks priority inasmuch as many scientists had previously implicitly or explicitly asserted that a quantitative or qualitative prediction that is invariant with respect to changes in underlying assumptions or parameter values was preferable (e.g. [32]).

For any scientist, the minimal practical implication of all of this back and forth is that it can help raise to a more conscious-level questions such as, what am I really saying when I claim that a model is more general, is the notion of generality meaningful, how is my claim about generality really grounded in the structure of the model, and should I believe more in a prediction made jointly by several models as compared to a prediction made only by one model? More common conscious consideration of these questions would unquestionably benefit the practice of biology. The third implication to arise from the debate concerns the need for change in the practice of biologists when working with models. In this context, it is typical to encounter the implicit or explicit attitude that a more general model is better. For example, Evans et al. [4] construed models that account for within-population dynamics to be more general and that the development of such models should be a main goal of ecology (when compared with continued use of simpler models that do not fully account for such dynamics). Inasmuch as one assesses generality appropriately (e.g. by explicit comparison, see §4), the search for general models may be useful. After all, we wish for many reasons to understand the broad patterns in nature and one can at least hope that their explanation will come most easily from a few widely applicable models (as opposed to a much larger assemblage of less general models).

However, in the development of general models, one must do more than just be comparative. Care is required in how one states any given model since this can influence the outcome of model comparisons. As Orzack & Sober noted [11, p. 535], consider a statement of the Hardy–Weinberg model as ‘If no evolutionary forces are present, then the genotypic frequencies are \(p^2, 2pq, \) and \(q^2\), where \(p\) and \(q\) are allele frequencies at a single bi-allelic locus. By comparison with a model that accounts for, say, the influence of natural selection, the Hardy–Weinberg model appears to be less general than the latter model as there are few populations that are not influenced by natural selection. Contrast this with the conclusion that one reaches when the assessment of generality involves a statement of the Hardy–Weinberg model as ‘If genotypic frequencies depart from \(p^2, 2pq, \) and \(q^2\), then some evolutionary forces are present’. Since genotypic frequencies in real populations typically depart somewhat from the expected proportions (possibly owing to the influence of natural selection), the Hardy–Weinberg model could be judged more realistic than a model that does not invoke natural selection.

Many biologists might say ‘these examples involve judging model generality based upon how the model is described but I judge the generality of a model purely on the basis of what processes and effects it does and does not represent mathematically. Surely, this is unambiguous.’ Perhaps, but one might reasonably expect then to fully dispense with language when assessing models and their value. It is inconsistent on the one hand to, say, count model terms in order to assess model generality (thereby avoiding the ambiguity of verbal assessment) and on the other hand to attach importance to the word ‘generality’.

The best alternative is to assess models purely on the basis of performance (say, how many cases for which it yields qualitatively or quantitatively accurate predictions), without an a priori overlay of value attached to it in the form of a label ‘general’ (i.e. good) or ‘not general’ (i.e. bad). Why are models labelled at all? It is easy to see that the tendency to do so is at least partially rooted in the desires and ambitions of an individual scientist who wishes to persuade other scientists to agree with the approach taken. If a word helps do this, so much the better. Persuasion is not problematic. What is problematic is when model assessments are poorly drawn and portrayed as inhering in the model itself as opposed to being at least partially determined by the social context in which we do science (see [33] for insights into the scientific and social contexts that influence the formulation and the assessment of scientific models).

Ultimately, it is important in this context to remember the human dimension of science. In searching for answers to real-world problems, we need all the help we can get—at least I do! If progress occurs, i.e. a particular approach works better, it is of no consequence that one might have been persuaded to try the approach simply because it was described as, say, more general. It is also of no consequence that the claim of generality may lack coherence. What is of consequence is that we judge the value of models as biological explanations in terms of their success at explaining patterns in nature. Even the best classification of this or that model as being general, realistic or precise, while perhaps not without some value, is vastly less important.

6. FUTURE CHALLENGES

I have described above how some of the philosophy of biology literature, whether it has been penned by philosophers or biologists, has real-world normative lessons for biologists that can be applied in analyses ranging from those at the level of cell to those at the level of ecosystem. Despite this, it is easy to be impatient (or worse) with the philosophy of biology as a formal discipline inasmuch as too much of the literature is detached from an understanding of the
practices of biologists and does not address questions of practice. (Some of this detachment possibly reflects a decision by philosophers that the discipline owes nothing in return to biology.) As a biologist, I favour the development of a philosophy of biology whose primary goal is to substantially inform and guide the practice of biology, so as to make it better (see [27] for a related discussion, as well as a presentation of additional topics in ecology that will probably benefit from philosophical scrutiny). In keeping with this goal, I now outline two challenges, which if met by biologists and philosophers of biology would immensely advance the practice of biology.

The first challenge in this regard is to understand how to analyse biological systems that encompass a wide range of phenomena. In the analysis of, say, ecosystem dynamics, biologists face practical and conceptual challenges in regard to how many levels of organization and causal processes to include. For example, many traditional analyses of energy flow in ecosystems have paid no attention to within-population dynamics. More recently, there are claims that the inclusion of such dynamics is necessary and sufficient to explain and predict dynamics at higher levels. Here, the focus is on the so-called ‘individual-based’ or ‘agent-based’ models [34–38]. In a conceptual sense, the important questions to be answered about the necessity and meaning of these new approaches at least partially come down to questions as to the nature and the extent of emergent properties and synergy in hierarchical causal systems (see [39,40] and references therein). Here, the philosophy of biology has much potential to contribute to a practical understanding of whether and how to include lower level processes (e.g. within-population dynamics) when analysing higher level processes (e.g. ecosystem dynamics). What is not sufficient (cf. [34]) is a decision to include a lower-level process purely because it has a measurable causal influence at a higher level. For example, it is one thing to acknowledge that, say, individual demographic differences influence the abundance of a prey species, which in turn, ultimately influences the food web of the entire ecosystem. However, it is something different to conclude that such demographic differences must be included in an analysis at a higher hierarchical level (e.g. the dynamics of the food web). In some way, every model is based on a decision to ignore some hierarchical levels. After all, we know that complex dynamics at the cellular-, tissue- and organ-level influence individual demographic differences, which in turn influence species’ abundances, and thereby influence ecosystem dynamics. But it is not credible (much less feasible) that we would create a model of ecosystem dynamics that was explicitly grounded in the metabolic dynamics of the cell. We need black boxes, which hide underlying details; the subtle question is where and when to use them. This is a golden opportunity for joint work by philosophers and biologists, one that could provide substantial normative insights. For example, Orzack & Sober [41] presented a practical framework to determine whether one can use a black box hiding the genetic and developmental details of a trait when analysing hypotheses about the power of natural selection to influence its evolution.

More generally, can consistent and biologically meaningful rules be found to determine the ‘location’ of black boxes? At present, most have locations that seem arbitrary, even if the ‘construction’ of the black box is exemplary. For example, Ludwig & Walters [42] showed by numerical analysis of a fisheries model that locating a black box around the age-structure of a harvested species (so as to ignore it) results in equivalent or better estimates of the optimal harvest. This is an important finding, but it begs the question as to whether this black box is more justified than any other; for example, both the age-structured and the non-age-structured models are deterministic and so the black box around environmental variability is not examined. Why is one black box privileged and not another? At least part of the answer is likely to be found via the use of statistical tools such as the Akaike information criterion (AIC). This allows one to balance the trade-off between the fit of the model to the data and the simplicity of the explanatory model (see [43]). What this likely will imply in practice is that in the analysis of, say, ecosystem or landscape dynamics, ‘fully’ causal models, i.e. those in which processes at the individual level and upward are included, will not be models of choice when compared with those for which some processes are ‘censored’ (cf. [34]). Instead, ‘mixed’ models, which fully account for some processes and not so for others could well prove most useful. In fact, such an outcome may turn out to be especially welcome inasmuch as the superiority of mixed models is clearly a joint consequence of model structure and of available data, cf. Hilborn & Mangel [44] whose basic principle for selecting among ecological models (p. 37) embodies this contingency: ‘let the data tell you’. A given model’s superiority could well disappear as more or better data become available. This contingency invites a pragmatic flexibility as to where to invest one’s time and effort and avoids the need to imagine that one must make a ‘final’ choice of model, especially a choice guided by an a priori decision as to whether a particular model is, say, general enough to be considered seriously. This flexibility is seen more and more often in the analysis of complex systems. For example, Orzack et al. [45] found in their analysis of the dynamic and static expression of life-history traits in a long-lived seabird, that temporal variation has a stronger influence than does age. Yet, they also analysed the influence of age on trait expression while ignoring temporal variation, the goal being to explore the biology with the deterministic framework most people use, so as to allow the results to be compared with those of previous analyses. There is no goal here of identifying the single ‘best’ model. This is just one of thousands of recent analyses in which this kind of flexibility occurs; many have relied upon the approach to modelling presented by Burnham & Anderson [43].

The manifest challenges of understanding complex systems practically demand this kind of agnosticism. It is ironic in this regard to note that Levins’s trichotomy of models and his endorsement of a particular type of model are now invoked by some in an effort to spur on the development of more synthetic ecological models.
even though his claims may well have contributed to the movement away from synthetic ‘systems ecology’ models starting in the 1970s (see [31]).

As biologists, we currently traffic in mixed models. Most often this practice appears to be guided mainly by convenience and tradition (e.g. ignoring environmental stochasticity in many population-dynamic models) rather than an explicit belief that something is unimportant. This is simultaneously an essential attitude and a dangerous attitude. It is essential because it allows one to censor a model (ignore, say, within-population dynamics) in an effort to explore ecological processes at other levels. It is dangerous because it can be highly misleading, at least to the extent that one can readily identify deterministic dynamical systems whose important qualitative features differ from those of analogous stochastic dynamical systems. For example, compare the standard Hardy–Weinberg model with two alleles in which genotypic fitnesses have equal fixed values, which results in stable proportions of all three genotypes, with a model in which genotypic fitnesses are equal on average but vary over time, which can result in quasi-fixation of an allele such that only one genotype remains (see Karlin & Lieberman [46]). Similarly, compare a deterministic density-independent population growth model when the logarithmic growth rate is zero, thereby guaranteeing population persistence, with a stochastic density-independent model when the average logarithmic growth rate is zero, thereby guaranteeing population extinction (see Orzack [47]).

Despite this potential for being misled by censoring, one can only applaud efforts to create a standard ecological modelling framework that allows for censoring, the goal being the development of rules for the inclusion or exclusion of this aspect of biological reality that are more science than ‘art’. Consider the remarkable presentation of Link et al. [48, pp. 62–64] of ‘best practice approaches’ for modelling ecosystems. Here, we see recommendations ranging from those that concern model structure (e.g. only explicitly represent primary productivity and nutrient cycling ‘when bottom-up forces or lower trophic levels are of key concern’) to those that concern model validation, such as

For dynamic models, best practices [sic] is to fit to as much data as possible using appropriate likelihood structures, while being clear about both potential biases arising from fixing parameters, as well as fully reporting error ranges resulting from freeing parameters. In case of fixing parameters, additional sensitivity analyses (e.g. resampling, Monte Carlo routines) should be used to assess model sensitivity to the assumptions. An important component is using results of sensitivity analyses to guide future data collections and the continuation of critical time series.

This is not the strategy of model building of 1966; in fact, this is a framework to create ecosystem models that are more general, more realistic and more precise than current models. It is not even the strategy of model building of 1996 in that it is predicated upon the availability of computational resources that were not available 15 years ago. The main point is that such an effort provides a perfect opportunity for philosophers and biologists to jointly refine normative guidelines in at least somewhat of a principled manner. Only time will tell if this happens and if operative guidelines actually result in better understanding of ecosystem dynamics. If so, a new standard will be set with respect to improving the process of modelling, as opposed to its content. Ultimately, it is likely that models that successfully predict important static and dynamic aspects of hierarchical systems like ecosystems will be causally fuzzy inasmuch as they will be mixed models (as defined above in this section). The fundamental justification for this is that such models work, i.e. they successfully predict what happens without necessarily telling us why it happens.

Sometimes, a different kind of justification of a mixed model is offered. For example, this or that deterministic ecological model is said to provide a ‘mean-field approximation’ to an otherwise identical stochastic model. Although this can be highly problematic for the reason just mentioned, it at least has the virtue of being an explicit acknowledgement of the modelling gambit that has been undertaken. Here, a particular kind of simplified causal process is assumed to operate in the black box (e.g. the influence of temporal variability of prey body size on the growth rate of its predator can be ignored and we assume that this influence can be adequately represented by the influence of the average prey body size on the predator growth rate). This may well work some of the time, but there is a need to be more aware of what is actually assumed in terms of biological processes. When ‘mean-field approximations’ are invoked in biology, they appear to be invoked as a matter of (arbitrary) tradition and there is a need to develop more standard criteria for their use. At least some useful starting points for this endeavour may be gleaned from solid-state physics, where the concept of the mean-field approximation arose. For example, one finds explicit dimensional criteria for assessing whether a particular mean-field approximation will or will not qualitatively describe the important features of a phase transition (see [49]). Perhaps, similar dimensional criteria can be developed for ecological models, so that guidelines for the inclusion and exclusion of simplified causal processes (like those of Link et al. [48]) will be more principled, as opposed to being determined more arbitrarily.

The second challenge for biologists and philosophers of biology is overcoming the cultural and institutional barriers to effective communication and collaboration. At present, many of the ‘live’ issues in the philosophy of biology are based on an outdated knowledge of biology. The outdated nature of some of the debates can be illustrated in the context of Levins’s claims about model building. For the most part, this debate in the philosophy of biology has the tacit assumption that Levins’s claims are descriptive of and relevant to current practice, say, in population biology. (This assumption is likely partially based on the belief that some implicit endorsement of the endeavour of philosophy of biology arises from referencing the claims of a distinguished biologist.) However, for better or for worse, much current development and use of models in population biology are far removed from the idealization in the philosophy of biology literature of Levins’s claims, which has at its core the notion that typically a
scientist strives to create a single explanation for the biological system under investigation. (This interpretation is consistent with the focus in Levin’s paper on a typology of models as a tool for classifying individual models, and with his endorsement of the goal of generating one ‘robust’ prediction from multiple models; on the other hand, he does allude to a satisfactory theory as being one that ‘is usually a cluster of models’ [10, p. 431].)

However, much of the current research in population biology bears little resemblance to this ‘single’ model or prediction idealization. As noted above, there is often no evident goal of deriving the best model (although this does occur sometimes); instead, given scientists often investigate a system with multiple models, with no attention paid to their reconciliation, even within the same publication (e.g. [45], and many others). It is this research microcosm that philosophers of biology ignore at the peril of irrelevance to the practices of biologists. At present, most approaches by philosophers of biology to understanding investigatory pluralism conceive of it as applying across scientists, as opposed to applying within a scientist within the same publication. This microcosm has some overlap with that described by Levins, but it engenders a substantial number of new conceptual and normative issues that philosophers of biology could do well to address, especially if they wish to be normatively relevant.

How might improved understanding of current biology come about? Some perspective can be gained by considering the history of philosophy of biology as a discipline. Although some core approaches and concepts date back many decades, the field has largely come together only in the past 30 years. As a result, standards of training lack uniformity, especially as judged over cohorts, with earlier cohorts of still active researchers (e.g. trained in the 1970s or earlier) having less formal training in biology (although many have made important contributions), whereas more recent cohorts often have research in biology as part of their degree programme. Time will tell whether this trend towards research and training in biology will result in a philosophy of biology that is more directly connected with current biology. At least the trend is encouraging, as is the increase in the one-on-one collaborations between biologists and philosophers and in the appointment of philosophers of biology to biology departments. On the other hand, philosophers of biology are rarely seen at professional meetings of biologists and most of their publications are not in biological journals.

Of course, the lack of connection between philosophers of biology and some important current developments in biology has its mirror image among biologists. Most biologists are also disconnected from some of the important understanding that one can gain from philosophy and from the philosophy of biology (see §3). Consider two examples of this disconnection. The first is the distinction often made by biologists between a ‘statistical’ model and a ‘biologically causal’ model. The former is often said to or implied to describe a relationship between two quantities, with no commitment being made as to causation; there often appears to be an implicit judgement that such models are not ‘real’ biology. The ‘biologically causal’ model is viewed as preferred. Although the conception of an acausal model has some formal meaning, in fact, the distinction between statistical and biological models usually detracts from the understanding to be gained from the biology at hand. For example, consider the estimation of a set of age-dependent survival rates using logistic regression, which relies upon statistical concepts and theory having no inherent biological content [50]. It is one thing to believe that such an analysis, with its estimate of a baseline survival rate (as embodied in an intercept term in the fitted regression model) and an age-dependent contribution to the survival rate (as embodied in the linear coefficient of age in the model), does not fully represent the biology connecting age and survival rate. This must be true. It is something quite different to assume that the use of such an approach necessarily divorces the results from describing biological causation. Instead, this approach can be better viewed as a hypothesis as to the location of a black box when attempting to understand causation in a biological system. The presumption is that, say, the linear coefficient of age embodies the net causal influence of age. In addition, since all models contain some black boxes (there is some philosophy of science literature on this topic, see [51]), there is no hard and fast distinction between a statistical model and a causal, biological one (cf. [52,53, p. 21]). Although this may appear to be an innocuous shift, it is profound in terms of practice. Perhaps of most importance in this regard is that it licenses the need to take seriously any model, and to not automatically dismiss a type of model as inferior because it is does not include ‘all’ of the biology. Instead, predictive performance is what matters.

The practice by biologists of judging model performance often involves a second kind of disconnection between philosophy and biology. In popular culture, it is a commonplace to describe scientists as comparing the predictions of models to data; it is one of the canonical activities that we associate with the endeavour of doing science. Yet, when it is done, it is very often done in such an inscrutable manner that the ‘scientific’ criteria by which model performance is judged are obscure. This obscurity is startlingly common if you look for it. The inferential confusion it engenders was illustrated by Orzack [54, p. 486] who presented published examples of conclusions for and against optimality of a sex ratio behaviour that contained no accessible basis for the judgement of the fit between data and predictions. All (or none) of the specific claims may be correct, but it is clear (without being orthodox) that none of the claims was scientific in the sense that they are based on a transparent recipe for judging model performance; the criteria for model confirmation were left unstated. A transparent and practical framework for grounding claims like these for or against optimality on the results of qualitative and quantitative comparisons between data and the predictions of optimality models was outlined in Orzack et al. [55] and further discussed in Orzack & Sober [41].

Vague practice when assessing model performance has had a special constituency in biology, one licensed by the often-heard attitude that ‘nature is complex’, which in turn appears to license the attitude that any prediction ‘near’ the data serves to confirm the validity of the model.

The ongoing popularity of this problematic practice is even more startling inasmuch as the typical...
quantitative abilities and judgements of biologists have increased substantially as compared with even the 1990s. Consider the recent very quantitatively sophisticated analysis of biomass dynamics in forest ecosystems by Antonarakis et al. [56]. Their goal was to understand whether various forms of remote sensing can supply enough information to accurately predict biomass dynamics. The two remote-sensing models used are censored in different ways inasmuch as the underlying remote-sensing techniques differ in what details of forest structure they record and, as a result, they generate different predictions. These predictions can be compared with data, which in this case arise from a complex simulation model (see [56, p. 1124]). However, despite the quantitative sophistication needed to carry out this research, their assessments of model performance ultimately amount to no more than visual inspection without any elucidation of what criteria underlie judgement of model success or failure. They write [56, p. 1127] in regard to the relationship between model predictions and simulated data shown in their fig. 6 that ‘Both forms of remote-sensing initializations have basal area and [aboveground biomass] diameter distributions that are close to the size distributions of the forest-inventory initialization’.

Visual inspections like this can be highly subjective and are known to be strongly influenced by the changes of the scale and magnification of the presentation (with smaller magnification often leading to improved assessment of performance and vice-versa; see Cleveland & McGill [57–61]). In fact, subjectivity in and of itself is not necessarily problematic; what is problematic is that each person’s subjective recipe for judging model performance is private.

In the case of Antonarakis et al. [56], ‘close’ is not defined and so one is left wondering how one might proceed if doing a similar analysis. How ‘close’ is close enough to merit acceptance of the model? How ‘far’ is far enough to merit rejection of the model? It is telling that there are readily available statistical techniques that would allow one to compare the predicted and observed diameter distributions. For example, one might choose a goodness-of-fit test (since the observations are based on simulations, one would choose some arbitrary sample size made known to the reader). In this way, at least an accessible recipe for analysis is ultimately amount to no more than visual inspection without any elucidation of what criteria underlie judgement of model success or failure. They write [56, p. 1127] in regard to the relationship between model predictions and simulated data shown in their fig. 6 that ‘Both forms of remote-sensing initializations have basal area and [aboveground biomass] diameter distributions that are close to the size distributions of the forest-inventory initialization’. Visual inspections like this can be highly subjective and are known to be strongly influenced by the changes of the scale and magnification of the presentation (with smaller magnification often leading to improved assessment of performance and vice-versa; see Cleveland & McGill [57–61]). In fact, subjectivity in and of itself is not necessarily problematic; what is problematic is that each person’s subjective recipe for judging model performance is private.

In the case of Antonarakis et al. [56], ‘close’ is not defined and so one is left wondering how one might proceed if doing a similar analysis. How ‘close’ is close enough to merit acceptance of the model? How ‘far’ is far enough to merit rejection of the model? It is telling that there are readily available statistical techniques that would allow one to compare the predicted and observed diameter distributions. For example, one might choose a goodness-of-fit test (since the observations are based on simulations, one would choose some arbitrary sample size made known to the reader). In this way, at least an accessible recipe for analysis is available, so that subsequent investigators can carry out comparable analyses. Without such a standard, one is left clueless as to the basis for conclusions regarding model performance. This gap detracts from an otherwise exemplary study like that of Antonarakis et al. [56]. This gap is very common in the literature.

A common reaction to a talk like this for quantitative testing of biological models is that it is asking too much of such models. Yet, science is full of models of complex phenomena whose performance is judged solely by quantitative means (often quite positively); examples include global climate models (e.g. Thorne et al. [62], fig. 7). The capriciosity of the attitude that it is asking too much of most biological models to yield accurate predictions is underscored by the longstanding practice of expecting accuracy for some biological models; a typical example is the Hardy–Weinberg model, which is routinely expected (correctly) to yield quantitatively accurate predictions as to genotypic proportions, despite the ‘complexity’ of influences on natural populations and despite the simplicity of the model.

The main point here is not that models in biology must yield accurate predictions in order to avoid being deemed as failures; instead, the main point is that models be tested quantitatively (and qualitatively) in such a way that conclusions about model performance be more science than art; the criteria for model assessment must be public, not private. The negative consequences of unstated criteria for model success are evident. For example, the literature in population biology is rife with conflicting claims about important issues, such as the power of natural selection to influence evolution in natural populations; one central reason for this is a lack of normative standards with respect to how to assess model predictions, as opposed to being a matter of ambiguity in the data [41]. Normative guidance by biologists and by philosophers of biology in regard to standards and practices of assessing the performance of models is sorely needed; there is little relevant work by philosophers of biology (see [63,64] for related discussions). Such guidance would be invaluable.

7. WHITHER PHILOSOPHY OF BIOLOGY?

The goal of improving standards of practice by biologists tasks biologists and philosophers of biology. Although no portion of this task should be the special provenance of philosophers, I am tempted to think that philosophers of biology will contribute most to achieving this goal by gaining expertise in statistics, in the psychology of visual perception and in the psychology and pedagogy of modelling. (These are apart from training in biology, see §6.) All of these will help reveal the way in which biologists actually form and use models. Such understanding is the first step towards normative guidance. The tandem of training in biology and of training in statistics, psychology and pedagogy is essential. The former has the unquestionable benefit of improved biological understanding, but the potential danger of making unconscious the practices of, say, model assessment that are most in need of reform; after all, training of graduate students is most often training in the field as it is, as opposed to what it should be. The training in psychology and pedagogy is essential so as to develop understanding of how scientists really treat models and their predictions and how scientists learn to do what they do. Ideally, this kind of understanding will make more apparent what practices are in need of reform and how to do this.

This is not the philosophy of biology of 2011. Even the current practitioners of philosophy of biology operate mainly with the ‘pure’ analytical mindset that is the trademark of philosophy. What is left out is an ‘applied’ connection with psychology and pedagogy as they relate to the use of models by biologists. For example, to my knowledge, none of the literature on the ways in which human perception influences model assessments (see §6) has been referenced in the philosophy of biology debates about modelling.
In addition, to my knowledge, in these debates, there have been no connections made to the substantial and relevant literature in psychology and pedagogy that concerns how scientists learn and actually practice modelling. Consider the edited volume by Lesh et al. [65], which contains numerous relevant contributions concerning the actual processes and practices of modelling in various disciplines and how they arise. (This volume is the published proceedings of the 2007 meeting of the International Community of Teachers of Mathematical Modelling and Applications; the focus of their 2011 meeting was ‘Mathematical modelling: connecting to practice—teaching practice and the practice of applied mathematicians’.) Philosophers of biology (and biologists) interested in modelling ignore this literature at their peril. The disconnection here speaks partially to the current separation between ‘high-status’ academic disciplines such as biology and philosophy, which are typically part of the faculty of arts and sciences in many universities, when compared with the ‘low-status’ study of education (which is often relegated to a separate faculty): out of sight and out of mind. Of course, this separation has not been and need not be forever; philosophers of biology need look only as far as John Dewey to find a philosopher who was simultaneously engaged with logic, practice and pedagogy. By similarly engaging with these, philosophers of biology can create a new philosophy of biology, one that is truly used by biologists as a guide to practice.

If the discipline of the philosophy of biology, whether it be created by philosophers or biologists or both, can contribute to the creation of normative guidance that is truly synthetic and engages with the real doing of biology, it will be a tremendous boon to our efforts to understand scientifically the consequences of global warming, habitat loss, resource loss and population growth. If the philosophy of biology is truly ever to have any use, here is the place and now is the time.

I thank T. Benton, J. Clark, M. Evans, B. McLoone and an anonymous reviewer for comments and criticisms, and T. Benton and M. Evans for the wonderful opportunity to participate in this Royal Society conference. Funding was provided by NICHD R03HD055685-01A2 and the National Academies Keck Futures Initiative.

REFERENCES


Is philosophy of biology useful?


