View metadata, citation and similar papers at core.ac.uk
brought to you by CORE
provided by Philsci-Archive

Model Organisms are not (Theoretical) Models

Arnon Levy and Adrian Currie

Forthcoming in *The British Journal for the Philosophy of Science.*

Abstract

Many biological investigations are organized around a small group of species, often referred to as "model organisms", such as the fruit fly *Drosophila melanogaster.* The terms "model" and "modeling" also occur in biology in association with mathematical and mechanistic theorizing, as in the Lotka-Volterra model of predator-prey dynamics. What is the relation between theoretical models and model organisms? Are these models in the same sense? We offer an account on which the two practices are shown to have different epistemic characters. Theoretical modeling is grounded in explicit and known analogies between model and target. By contrast, inferences from model organisms are empirical extrapolations. Often such extrapolation is based on shared ancestry, sometimes in conjunction with other empirical information. One implication is that such inferences are unique to biology, whereas theoretical models are common across many disciplines. We close by discussing the diversity of uses to which model organisms are put, suggesting how these relate to our overall account.

1. Introduction
2. Volterra and Theoretical Modeling
3. Drosophila as a model organism
4. Generalizing from work on a model organisms
5. Phylogenetic inference and model organisms
6. Further roles of model organisms

   6.1 Preparative experimentation.

   6.2. Model organisms as paradigms

6.3. Model organisms as theoretical models.

6.4. Inspiration for engineers

6.5. Anchoring a research community.

7. Conclusion

1. Introduction

Many biological investigations are organized around a small group of species, often referred to as "model organisms", such as the bacterium *Escherichia coli*, the fruit fly *Drosophila melanogaster* and the house mouse, *Mus musculus*. Research employing these organisms has led to key discoveries: basic mechanisms of heredity were discovered in *Drosophila*, simple but powerful forms of gene regulation were first understood in *E. coli*, and much of our knowledge about cancer and metabolic diseases comes from work on mice. The terms "model" and "modeling" also occur in biology in association with mathematical and mechanistic theorizing, as in the Lotka-Volterra model of predator-prey dynamics, the Hodgkin-Huxley model of the action potential and the French Flag model of cellular differentiation. Let us call these *theoretical models* to distinguish them from model organisms.

What is the relation between theoretical models and model organisms? Broadly construed, a model is a cognitive stand in: instead of investigating the phenomenon directly, one studies an easier to handle alternative. In this loose sense, both *Drosophila* and the Lotka-Volterra equations serve as models. However, we shall argue that model organisms and theoretical models differ substantially in epistemic character. Model organisms serve as samples from, or specimens of, a wider class. In contrast, theoretical models, as the name suggests, are constructs that serve as theoretical analogs of their targets.

A number of recent authors appear to suggest, *contra* this view, that model organism research should be assimilated to theoretical modeling. Sometimes the suggestion is implicit. For instance, Ankeny and Leonelli ([2011]) state that model organisms can be understood within the "models as mediators" framework (Morgan and Morrison, [1999]), developed primarily to handle theoretical modeling. Similarly, in their Stanford Encyclopedia of Philosophy entry "Models in Science", Frigg and Hartman ([2006]) enumerate various kinds of concrete theoretical models (such as scale models in engineering and electric circuit models in neurobiology). They refer to model organisms as a "more cutting edge" instance in this category. Michael Weisberg argues more explicitly that model organisms are a kind of concrete theoretical model, differing only in that they are not artificially constructed ([2013], §2.5). This suggests that the status of model organisms is worth hashing out. That said, we accept many elements of what the aforementioned authors say about model organisms, and we do not wish to engage in a polemic. Therefore, we shall largely be concerned with our positive account.

Our focus is the epistemic features in virtue of which organisms (on the one hand) and theoretical constructs (on the other hand) serve as models. We aim to elucidate the basis upon which biologists make inferences from results obtained in a model to a different, typically broader class of phenomena. To this end, we set aside several issues surrounding models. For one thing, our discussion doesn't touch on ontological or semantic questions, such as what models are or how they represent. For another thing, although we rely some historical literature to support our claims, we do not offer a historical or sociological story *per se*. The aim is to account for the justificatory structure underlying the inferential move from models to targets. Moreover, our argument does not depend on the usage of 'model', 'modeling' and kindred terms. We use the labels 'theoretical models' and 'model organisms' because we find them appropriate and we think they are consistent with some, but perhaps

not all, scientific and philosophical usage. But nothing of substance turns on how biologists, or philosophers, use the terms in question.

The paper proceeds as follows. In section 2 we lay out a view of theoretical modeling. Section 3 turns to model organisms, focusing on *Drosophila.* In section 4 we compare the two, arguing that they involve different epistemic practices. With this basic conception on the table, we refine the discussion in two ways. First, in section 5, we look in more depth at the role of phylogeny in inferences from model organisms, showing that it is a form of the comparative method more generally, and as such an epistemic resource that is *unique* to biology (or near enough). Section 6 expands on the picture by sketching several further roles of model organisms, linking them to the earlier ideas.

2. Volterra and Theoretical Modeling

To characterize theoretical modeling we will first look at an example, the Lotka-Volterra model of predator-prey dynamics, and then offer a more general discussion. This example has received significant attention in recent philosophy of biology. That, in part, is why we have picked it: at the risk of being unoriginal, we focus on a familiar and relatively uncontroversial case.

The applied mathematician Vito Volterra's interest in the dynamics of predator-prey systems was sparked by empirical observations made by his son-in-law, Umberto D'Ancona, a marine biologist. During World War I, fishing in the Adriatic Sea all but ceased. D'Ancona discovered that, curiously, the lack of fishing seemed to advantage predators: right after the Great War their numbers were proportionately higher than they were before it. Volterra analyzed the situation mathematically, making a number of simplifying assumptions, e.g. that fish populations were well-mixed so that encounters among individuals were random. He stipulated that absent predation, prey populations

would grow without limit. And he treated predators as single-mindedly pursuing one kind of prey. However, Volterra was careful to retain several key features of real predator-prey relations, including, importantly, '[That] the proportional rate of increase of the [prey] species diminishes as the number of individuals of the [predator] species increases, while augmentation of the predator species increases with the increase of the number of individuals of the [prey] species' (Volterra, [1926], 558). This property is nowadays known as negative coupling, and it is the core of predator-prey relations. Volterra formalized this setup, producing the following set of ordinary differential equations:[i]

(1) $\frac{dV}{dt} = rV - (aV)P$

(2) $\frac{dP}{dt} = b(aV)P - mP$

Equation (1) tracks the abundance of prey (*V*): the first term represents the prey's growth rate, and the second the rate at which prey are captured by predators. Equation (2) tracks the abundance of predators (*P*): the first term represents the rate at which prey is "converted" into new predators, while the second the rate of predator mortality. Volterra's analysis indicated that predator and prey populations exhibit distinctive, out-of-phase oscillations. Most significantly, it showed that killing off both predators and prey at a rate proportional to their abundance increases predator population while lowering prey. (In later work Volterra dubbed this "the Law of the Disturbance of the Averages.") This accorded with D'Ancona's observations: at the time, fishing removed both predator and prey at a rate proportional to their abundance. Considering the basic structural similarity between his model, especially the effect of fishing on a pair of negatively coupled populations, Volterra

concluded that the model's results explained D'Ancona's observations. It showed that '…closure of the fishery was a form of 'protection', under which the voracious fishes were much the better and prospered accordingly, but the ordinary food-fishes, on which these are accustomed to prey, were worse off than before.' (Ibid, 559).

Thus, Volterra constructed an idealized scenario, retaining some key structural features of real world fisheries. He then showed that the constructed scenario exhibited the phenomenon that was of interest in the real world case. The match between the real world phenomenon and the model, he argued, provides grounds for taking the model to capture the key goings-on, thus explaining D'ancona's original observation. Volterra furthermore believed, as he stated explicitly in several places, that any system that exhibits these basic structural features, would be amenable to a similar analysis.

Let us describe the general category of theoretical models – a term we reserve for work akin to Volterra's. In this, we partly rely on the picture of modeling proposed by Peter Godfrey-Smith ([2006]) and Michael Weisberg ([2007], [2013]). They view modeling as an indirect, surrogative method of representation and analysis: in modeling, a scientist learns about a target system not by studying it directly, but by constructing a modified version of it, which retains some features while simplifying others. The model is then analyzed, and results about its behavior are obtained. [ii] Armed with an understanding of the simplified surrogate, the modeler then assesses whether the retained features suffice to license an application of the model – perhaps only in part, or only in some contexts – to the real-world target. Volterra's work is a paradigmatic example of this mode of theorizing. Instead of describing fish populations directly, he constructed a mathematical setup, and showed that it exhibited stable oscillations, obeying a "law of disturbance of the averages." He then reasoned that the model retained enough of the core features of his real-world target (Adriatic fisheries), that it could be treated as an explanation of the observations made by

D'Ancona. Thus, Volterra performed a kind of analogical reasoning, moving from an analysis of a mathematical construct, the model system, to conclusions about a different sort of thing: a real-world target system.

What we are calling the model system can be an actual concrete object, but is more often a set of mathematical equations or a hypothetical mechanism. In either case, the model is a construct insofar as its properties are either wholly stipulated or specified so as to represent some target. Its elements and their arrangement are chosen by the modeler, who makes simplifying assumptions and idealizations in the process. One upshot of the model's constructed nature is the modeler's intimate knowledge of, and high degree of control over, its makeup. This makes its study easy compared to the target system. Furthermore, intimate acquaintance with the model guides the modeler in assessing model-target inferences. Volterra, for instance, knew that negative coupling was a crucial aspect of his model, and that was a key reason why he focused on this feature when applying his theoretical findings.

In sum, theoretical modeling involves a mathematical or mechanistic construct that serves as an analog of the target. The modeler analyzes the model and then assesses whether the target is sufficiently analogous to it. If successful, this analysis licenses conclusions about the target, on the basis of results concerning the model.

3. Drosophila as a model organism

Can this picture of theoretical modeling encompass model organisms? We hold that in key epistemic respects it cannot. Rather, we suggest that inferences from work on model organisms are empirical extrapolations, whereby biologists treat the organism as a representative specimen of a broader class.[iii] Our discussion proceeds in two steps. First, we describe the general features of model organism research, drawing on historical work. Then, in the next section, we look into the epistemic basis for model organism-based inferences.

We focus on *Drosophila* throughout, but also allude to other examples to highlight certain points.

The fruit fly *Drosophila melanogaster* is a central model organism in genetics and developmental biology. It rose to prominence in the first decades of the 20th century, through the work of Thomas Hunt Morgan and his group at Columbia University (later at Cal Tech). During the middle third of the 20th century its centrality to biological research waned somewhat, but as molecular genetics emerged, especially in the 1970s, the fruit fly came to re-occupy center stage (Keller, [1996]; Weber, [2007]). Both periods are relevant for our analysis. Work on *Drosophila* enabled the Morgan group to identify and characterize the phenomenon of genetic linkage and to develop chromosomal mapping. More generally it led to the articulation of, and lent initial support for, the chromosome theory as a mechanistic explanation for Mendel's rules and for the all-important exceptions to them. *Drosophila* was suited for work in the lab because its size and short generation time enabled the maintenance of large lab populations and allowed the observation of many cycles of reproduction and inheritance.[iv] Extra-biological reasons may have also played a role, such as the match between its seasonal life-cycle and the academic calendar (Kohler, [1994]). The fruit fly's centrality was cemented as further useful features were discovered, such as the giant chromosomes in its salivary gland.

Morgan and his students initially collected flies in the wild (i.e. the window sill or the backyard, as fruit flies live in close quarters with humans). But within a few years most flies were lab reared, as is true today. To enhance reproducibility and allow for easier comparisons, strains of interest were isolated, bred and standardized in the lab. In time, fruit fly genetics were deliberately modified to generate strains that were better suited for lab work – more viable, easier to score, simpler to cross etc. Morgan and his colleagues developed an array of experimental tools for working with *Drosophila*, from specialized

tubes and bottles to crossing schemes. Later workers perfected these tools, and expanded the toolkit. Nowadays virtually any gene in the fruit fly can be expressed at a specific stage and location, and gene expression patterns can be monitored in detail. More generally, the range and precision of available techniques for manipulating cellular and molecular structures in *Drosophila* (especially *melanogaster*) is greater than for almost all other species.

From the early days, work on *Drosophila* has been regarded as a basis for claims about other organisms. Morgan and his group treated their findings as applicable to a vast range of species – including moths, pigeons, cats, silkworms, rabbits, and several species of plants (e.g. Morgan et al. [1915]). They did not, in any text we know of, offer detailed statements about scope, but they make clear that they view results obtained in flies as indicative of the basic mechanisms of Mendelian heredity in a wide variety of organisms, perhaps all sexual species[v]. This judgment, with respect to *Drosophila* as well as other model organisms, is echoed by more recent biologists. As a recent genetics textbook puts it:

> The science of genetics discussed in this book is meant to provide an understanding of features of inheritance and development that are characteristic of organisms in general. Some of these features, especially at the molecular level, are true of all known living forms… [S]o we do not have to investigate the basic phenomena of genetics over and over again for every species. In fact, all the phenomena of genetics have been investigated by experiments on a small number of species, model organisms, whose genetic mechanisms are common either to all species or to a large group of related organisms. (Griffiths et al., [2008], p. 17)

To summarize: Model organisms begin their career by being collected from the wild, typically because they are easy to rear and convenient for the research at hand. Once brought into the lab, a model organism typically undergoes a process of genetic standardization and over time an intricate array of experimental tools and methods are developed. Finally, and (for our purposes) most importantly, results from experiments on model organisms tend to serve as bases for conclusions about *other organisms*; sometimes, as in the case of early studies in Drosophila, results are seen as very widely applicable – even as far as to all sexually reproducing organisms.

4. Generalizing from work on a model organisms

We now wish to argue that model organisms diverge epistemically from theoretical models. Theoretical models like Volterra's are idealized constructions, specified for analogy with a chosen target. Model organisms, in contrast, are drawn from a wild population. As we have noted, they typically undergo standardization and modification. But they are still treated by biologists, in most cases, not as artificial constructs but as members of the class of objects (i.e. organisms) under investigation (Weber, [2005]). The fact that an object is a member of a broader class, however, doesn't yet justify generalizing from it to other members of the class. How are findings from model organisms applied to other organisms? We consider three routes. One option is to look at the target organism directly, and investigate whether it has the feature first found in the model organism.[vi] This is a common practice in molecular genetics, where researchers try to ascertain whether a DNA sequence discovered in one organism occurs in other organisms and, if so, whether it exhibits the same activity. But here, the model organism isn't serving as a basis for *inference* about other organisms: it does not serve as an epistemic stand in, but as a guide for what to look for. Thus this is not a kind of modeling in the relevant sense.

Another, more model-like way in which a result from one organism may be generalized is via "circumstantial" evidence. Here one knows, with respect to some broad and/or partial features, that the model resembles some target range of organisms. From this, it is concluded that a *specific* finding in the model is likely to hold in the target range. For example, one of the most celebrated results in neuroscience is Alan Hodgkin and Andrew Huxley's discovery of the mechanism of the action potential ([1952]). In a nutshell, they showed that action potentials result from a specific chain of molecular events, wherein the neuron's permeability to ions of sodium and potassium rises and falls in turn. Hodgkin and Huxley's experimental work was done primarily in the giant axon of the common Atlantic squid (*Loligo paelleii*), a central model organism in neurophysiology. It had been previously shown that sodium and potassium have similar effects in neurons from various other organisms – including cuttlefish, frogs, mammalian hearts, algae and crustaceans. These results did not *prove* that action potentials worked similarly across these organisms, but, as Hodgkin and Huxley put it, '…the similarity of the effects of changing the concentrations of sodium and potassium on the resting and action potentials of many excitable tissues (Hodgkin, [1951]) suggests that the basic mechanism of conduction may be the same as implied by our equations…' ([1952], p. 542).[vii]

In this type of inference, the model organism fulfills a stand-in role of sorts. A coarse-grained uniformity across a range of organisms, coupled to a specific result from the model organism, are jointly taken to imply that the specific result from the model is likely to hold more broadly. Here, the model organism is treated as a specimen, and what we have called circumstantial evidence justifies treating it as representative of a broader class. In other words, such circumstantial evidence suggests that a certain range of organisms (including the model) is sufficiently uniform so that results obtained in the model can be generalized to the class as a whole.

A third way in which model organism generalizations are justified is via phylogeny. The basic rationale, to quote a central cell biology textbook, is that "because genes and gene functions have been so highly conserved throughout evolution, the study of less complex model organisms reveals critical information about similar genes and processes in humans." (Alberts et al., [2008], p. 556). This method is perhaps the most model-like of those discussed so far, and more importantly, it is distinctively biological. We expand on it, and on the broader method to which it belongs, in the next section. Here we provide a summary.

In a phylogeny-based generalization results from a species are extrapolated on the basis of evolutionary relatedness. This inference is guided by the assumption – or at least the hope – that creatures in the broader class, i.e. on the relevant portion of the phylogenetic tree, have retained the relevant features from their common ancestor. This inference also treats the model organism as a specimen, but the specimen's representativeness is justified via an assumption about the evolutionary history of the model and target. Information about relevant casual and behavioral similarities is replaced by an appeal to shared ancestry. The extent to which one can generalize on the basis of phylogeny is often difficult to ascertain, because of uncertainty both about relatedness and about the conservation of features. For these reasons, biologists often refrain from specifying the exact scope of generalizations from model organisms. It is usually safe to assume that the more basic a feature or mechanism is – in physiological and/or developmental terms – the less likely it is to change in the course of evolution[viii], and the more likely it is that results pertaining to it in one organism may generalize beyond it[ix]. This was the case with Morgan's work on key aspects of inheritance in sexual species. Alberts et al. make the point with respect to work on the cell-cycle in the nematode *Caenorhabditis elegans*:

> Although the worm has a body plan very different from our own, the conservation of biological mechanisms has been sufficient for the worm to be a model for many of the developmental and cell-biological processes that occur in the human body. Studies of the worm help us to understand, for example, the programs of cell division and cell death that determine the numbers of cells in the body—a topic of great importance in developmental biology and cancer research. ([2008], p. 37).

The two model-like inference patterns that we have discussed are a form of empirical generalization. In this respect, model organisms serve as bases for induction – specifically, they serve as bases for extrapolation from a specimen to a broader class. In both circumstantial evidence based inference and phylogeny based inference, the move from model to target is grounded in information pertaining to the representativeness of the specimen. Note that these methods are not mutually exclusive. Indeed they are commonly employed synergistically. This occurs when both circumstantial evidence and information about shared ancestry are available – and together these jointly support (and make specific) the projection from model organism to target.

In sum, we suggest that results from organisms can serve as bases of inference in one of two ways. The first involves an appeal to circumstantial evidence, so as to generalize to the likely applicability of the result in the range of cases for which there is such evidence. Alternatively (but not exclusively) the move to the target may be grounded in the phylogenetic relatedness of the model organism and the target range. These two forms of inference are broadly model-like. But they diverge in their *epistemic* roles from theoretical models. The type of stand in at issue is different. In theoretical modeling, model-target inferences are grounded in an explicit procedure of feature-matching. In model organism

work, the inference from model to target is mediated via indirect evidence about the similarity of members of the broader class of organisms to which both model and target belong. One kind of indirect evidence is what we have called circumstantial evidence, the other is shared phylogeny. This latter form of inference is distinctively biological, and we think it sets apart model organism work from other kinds of theoretical methods. We expand on this point in the next section.

5. Phylogenetic inference and model organisms

We believe model organism work is part of a distinctive class of biological inference strategies, known as the *comparative method*. While theoretical models are assessed for structural resemblance to real world targets, biologists engaging in the comparative method gain epistemic traction through ancestral relations. We sketch the comparative method and illustrate it, using an example that does not involve a model organism. Afterwards, we argue that model organism work has the same general form.

The Darwinian insight that all life is ancestrally related is at the heart of the comparative method. We can conceptualize ancestral relations between organismic traits as either *homologous* or *homoplastic*. Homologous relationships are those of common descent: the ancestor of the two lineages had that trait, and its descendants have inherited it. Homoplastic traits, by contrast, evolved independently – the common ancestor did not have the trait[x]. By comparing different lineages biologists can infer ancestral relationships (Sober [1988a]), frame and support adaptive explanations (Currie, [2012]; Griffiths, [1996]; Sansom, [2003]), infer unknown characters (see below), set molecular clocks (Ayala [2009]) and detect large-scale patterns in life's shape. A concrete case illustrates phylogenetic inferences.

The comparative method is often used to infer unknown traits. This is most prominent in paleontology, where the incompleteness of the fossil record necessitates reliance on inference from contemporary critters to access extinct lineages. Some extant animals present similar difficulties. The Colossal Squid *Mesonychoteuthis hamiltoni*, for instance, is both rare and lives in a high-pressure deep sea environment, making direct study next to impossible. In particular, its feeding behavior is a mystery. Biologists infer the behavior of colossal squid from organisms which they can access: that of closely related, smaller, and common squid living closer to the surface.

Mesonychoteunthis Hamiltoni has a fearsome reputation: the world's heaviest invertebrate at around half a ton, sporting the largest eyes and beak of any cephalopod. It is tempting to view them as dynamic, fast-moving, chase-and-kill predators. Rosa and Seibel ([2010]), however, argue that the Colossal Squid is not suited to the chase, and suggest that '… it is, rather, an ambush or sit-and-float predator that uses the hooks on its arms and tentacles to ensnare prey that unwittingly approach.' (p. 1376). We are not here concerned with the success or otherwise of their argument, but rather draw on their paper to illustrate phylogenetic inference. Rosa and Seibel's argument has two parts: first, they estimate the squid's metabolic rate in order to work out its daily prey requirement; second, they compare the Colossal Squid's daily prey requirement to that of other lineages which occupy a high-speed predatory niche. Each step involves a different use of the comparative method: the first is based on phylogeny, as in standard model organism work; the second infers between ecotypes. We focus on the first part as an example of a *non-model organism* application of the comparative method.

There are no live specimens of colossal squid available to directly measure metabolic rates. Fortunately, the *Cranchiidae* (Cranch) family includes smaller, more accessible lineages on which we *do* have metabolic information. Rosa and Seibel estimated the

metabolic rate of Colossal squid from measurements in Cranch squids of four size magnitudes. Relying on a general metabolic model they were able to estimate the rate of metabolism in Colossal Squid and compare it to other large squid such as the Giant (*Architeuthis*) and Jumbo (*Dosidicus gigas*). This suggested that the Colossal Squid requires *far less* food than other top predators of the southern oceans. This discrepancy led Rosa and Seibel to hypothesize a much more sedentary lifestyle for the Colossal Squid than both its smaller relatives and top predators in the sub-Antarctic oceans.

In inferring the metabolic rate and daily prey requirements of Colossal Squid, Rosa and Seibel perform a *phylogenetic inference*. The inference is in two steps. First, the trait of interest (in this case metabolic rate) is examined in one or more closely related lineages. Because *Cranchiidae* share a common ancestor, it is thought that traits held commonly among that clade were most likely also held by their common ancestor. The first step, then, is a retrodiction from contemporary lineages to the common ancestor of those lineages. The second step projects from the common ancestor to the target – in this case the Colossal Squid. It is thought that any trait held by a relatively recent ancestor, or a trait which is relatively entrenched, is likely to be retained in a contemporary lineage. By examining relatives, biologists postulate a regularity across the clade in question, maintained due to common descent. In Rosa & Seibel's case, the inference is also mediated via a metabolic model, but this is not always the case.

Figure 1 here

Figure 1 caption here

The colossal squid is not a model organism – and this is partly why we have chosen to discuss this example. The pattern of phylogenetic inference is, we submit, the same pattern

seen in standard model organism cases. Just as Morgan took *Drosophila* to be a representative sample of the basic genetics of sexual organisms, so Rosa and Seibel take accessible members of the *Cranchiidae* to be a good sample of that clade metabolically. In both cases the justification for the inference is not based on a direct comparison of known features, but rather on ancestry. The details – such as the kind of trait in question and the recency of the relevant common ancestor – differ, but the strategy is the same: the relatedness of the lineages licenses inferring from one to another, without the need to explicitly compare the underlying traits. Standard use of model organisms, then, is best understood as an application of phylogenetic inference.

Moreover, some criticisms of modern biology's reliance on model organisms can be understood in light of our discussion. For instance, Bolker and Raff ([1997]) argue that because model organisms are chosen, in part, on the basis of experimental tractability, they tend not to exhibit common features of the living world essential to its understanding, such as complex life cycles and certain forms of phenotypic plasticity. The claim, in essence, is that experimental considerations bias the choice of model organisms. Our discussion brings out the problem. Since model organisms serve as specimens, a bias in the criteria for specimen choice will affect the scope of consequent findings; they will apply only to a subset of extant lineages. Another kind of critique has been voiced by some microbiologists, who argue that extensive horizontal gene transfer (HGT), in model micro-organisms (especially bacteria such as *E. coli*) undermine generalizations across unicellular organisms. This is in contrast to inferences in multicellular creatures where HGT doesn't occur. Again, this dovetails with our discussion. To put it very briefly, HGT confounds attempts to identify bacterial lineages.[xi] For this reason, it poses serious challenges to the use of the comparative method, including inferences from model organisms.

In the next section we broaden the picture and consider additional functions played by model organisms. But before doing so let us note the place of phylogenetic inference in biology as a whole, as this reveals a further discrepancy between model organism work and theoretical models: while the latter practice exists in many parts of science, the former is particular to biology.

Phylogenetic inferences are made on the basis of *common causes*. A common cause explanation works on the assumption that a hypothesis about the past which unifies the most contemporary evidence is more likely than one which doesn't (Reichenbach [1956], Cleland [2011]). Contemporary philosophers have emphasized the role of common causes in historical inferences. Some (particularly Sober [1988b] and Tucker [2011]) argue that this inferential structure does not come for free: what *licenses* the assumption that contemporary events are likely to have common causes? After all, the world is a complex place, and it is not obvious that we should expect such uniformity.

Grounding common causes in phylogeny – where it can be established – meets this challenge. Morgan et al. had license to assume that most sexual organisms are genetically alike, and Rosa & Seibel have good reason to believe that metabolic rates are extrapolatable across Cranch squid. Both are justified by evolutionary theory. The high-fidelity cross-generational transmission of traits central to heredity leads us to expect a certain phylogenetic 'inertia' – often, a trait present in some past lineage will be inherited by its ancestors. Rates of inertia and the kind of traits involved will differ, and this will matter greatly to the kinds of inferences that common ancestry licenses, as well as their certitude (Sober [1988b]). But the underlying point still holds: shared ancestry serves as a basis for inferences as it is a common cause of traits across related taxa. These sorts of inferences have a special role in biology, because of their grounding in evolutionary theory[xii].

This justification, particular as it is to biological theory, stands in sharp contrast to the generality of the theoretical modeler's strategy: the use of analogical reasoning is widespread in biology but is, if anything, more common the physical sciences, in economics and a variety of other disciplines. Thus, phylogenetic inference, including model organism work, is distinct from theoretical modeling both in its justificatory structure and its domain.

6. Further roles of model organisms

Our focus in this paper is on the sense in which model organisms are models, i.e. in their epistemic stand-in role in biological practice. We have argued that this role is best understood in terms of empirical extrapolation, as a form of the comparative method. However, we do not wish to claim that the forgoing discussion is exhaustive: model organisms serve other important roles, epistemic and otherwise. Indeed we believe that properly understanding the work we have discussed requires situating it with respect to these other roles. That is the goal of this penultimate section.

6.1. Preparative experimentation. The importance of model organisms to biological research is due in large measure to their amenability to experiment. To some extent, this is because of natural biological properties, such as small size, short life-cycle and ease of adjustment to life in the lab. But many decades of work on model organisms have greatly contributed to their suitability to research, in providing crucial background information for designing and interpreting experiments, and ever more sophisticated methods of detection and analysis. Weber ([2005], §6.6) calls this "preparative experimentation". As he explains, "this kind of experimental work is not directly aimed at testing a specific hypothesis, nor do biologists necessarily need a guiding theory for conducting this kind of research. This does not mean that they do not need any theoretical knowledge. Clearly, developing experimental organisms and other research materials requires some knowledge of genetic mechanisms,

chemical properties of biomolecules such as DNA or protein, and so on." (Ibid, p. 174). As Weber shows, decades of classical genetic research on *Drosophila*, flowing from the work of Morgan and his colleges, provided knowledge about mechanisms of inheritance, but *also* material resources, such as methods for rearing, breeding, and genetically modifying the organism, strains with specific mutations that may be used as controls and/or as markers and cloned DNA fragments that serve as vectors and for detection purposes. These resources contribute to the entrenchment of a model organism within a research community, as they make future work on it easier and potentially more productive. To Weber's discussion we might also add that oftentimes, methods first developed in one model organism are exported to other experimental organisms. For instance, the UAS-GAL4 system, a powerful method for targeted gene expression first developed in *Drosophila* (Brand and Perrimon, 1993) has been extended for use in other organisms, including the frog *Xenopus laevis*, Zebrafish (*Danio rerio*) and mice.

Preparative experimentation isn't in itself model-like in character. It is empirical or methodological work aimed at facilitating future research.[xiii] However, it may contribute to the generalizability of model organisms in at least two ways. First, the more able scientists are to make discoveries about model organisms, the more material for potential for extrapolation there is. Secondly, when methods developed in a model organism are exported for use in other organisms this may not only enhance research in the organisms to which the method has been extended but also, at least in some cases, makes such work more comparable across species – partially controlling for differences in methodology – thus enhancing generalizability.

6.2. Model organisms as paradigms. Model organisms could be loosely described using Kuhn's notion of a *paradigm result*[xiv]. A paradigm result serves as an exemplary piece of science, guiding future researchers' expectations and standards of evaluation. It sets the

bar for best scientific practice. Thus, the early (and successful) effort in sequencing the model bacterium *E. coli* (Blattner et al., [1997]) was, among other things, a paradigm for future and ongoing sequencing projects. It helped set standards for what counted as sufficient genome coverage, and it facilitated the development of methods for parsing sequence data and organizing and presenting results.

This paradigm-result-role gives rise to methodological and epistemic standards. In a loose sense, paradigm results serve as a model for how to do science. This sense of 'model' may be quite different than our previous usage, so we employ it cautiously. It appears rather indirectly related to the role of model organisms in grounding biological generalizations.

6.3. Model organisms as theoretical models. We have distinguished between target-to-model inferences in theoretical modeling versus model organisms. The distinction concerns justificatory structure and not ontic character. For all we've said, there is nothing to block the use of an organism as a theoretical model. Recall that a theoretical model, as we use the term, serves as an analog or surrogate. Theoretical models often consist of mathematical equations, as in the Lotka-Volterra case. But models can be concrete objects, such as Watson and Crick's well-known wire-and-metal-sheets model of DNA. To serve as a model, a concrete object need not be manmade; it can be an organism.

Experimental evolution contains an important class of such cases. In Wade's ([1977]) study of the dynamics of group versus individual selection he subjected populations of flour beetles to different selection regimens, some of which favored group selection whereas others favored selection at the individual level. Or, more recently, Ratcliff et. al ([2012]) subjected yeast to selection pressures that favor increasing size, thus inducing the formation of many-celled clumps. This, they argue, amounts to the *de novo* evolution of multicelularity. In such cases it is clear that working with organisms isn't a means for generalization over related taxa. Wade did not take his results to apply in any special way to

beetles or insects, nor do Ratcliff et. al view their work as specially pertinent to yeast or to related microorganisms.[xv] Rather it is a form of theoretical modeling, where the model is a whole organism or even a population of organisms, which serve as a surrogate for wild organisms and populations, either in the past, over extended temporal or spatial scales, or in difficult-to-study locations and conditions.

To be sure, there are differences between modeling that utilizes concrete objects, organisms in particular, and mathematical or other abstract models. It might be said, for instance, that working with actual organisms provides results with a kind of "proof of concept" status, which mathematical theorizing cannot obtain (Crone & Molofksy [1999]; Odenbaugh [2006]). Perhaps so, but this is consistent with thinking that both kinds of cases are exercises in theoretical modeling. The broader and more important point is that a scientific study using a model which happens to be an organism doesn't suffice to make it a model-organism based study in the presently relevant sense. An organism can serve as a vehicle for theoretical modeling.

6.4. Inspiration for engineers. A role played by model organisms which is less internal to biology is guidance and inspiration for engineers.[xvi] For example, Ma and colleagues ([2013]) designed a flapping-wing insect-sized robot. Their design was inspired by studies on flight related morphology and behavior in flies, especially *Drosophila melanogaster*. It is possible to describe this sort of work in terms of modeling (Ma et al. state that their robot was '…modeled loosely on the morphology of flies.', Ibid, p. 603). Yet the role played by knowledge about *Drosophila* in this context is clearly different than the role it played in, say, the Morgan group's research. Most obviously, biologically inspired engineers need not aim for an understanding of the inspirational phenomenon, but rather look to construct an artificial device. That said, there can be a special role for model organisms in this context, because of the vast knowledge that exists about their biology.

Moreover, the construction of engineering models may feed back into our understanding of animals–constructing robots to mimic animal behavior, for instance.

6.5. Anchoring a research community. Finally, stepping outside of our epistemological focus, model organisms typically serve to anchor a scientific community. Several authors, both philosophers (e.g. Ankney and Leonelli, [2011]) as well as scientists (Griffiths et al., [2008], 759) view this as a central aspect distinguishing model organisms from other experimental organisms. It has been suggested that a community devoted to studying a particular model organism will tend to converge more closely on methods and standards of evaluation, and develop more extensive infrastructure, such as stock centers and databases. Such a community will also tend to follow specific, often tighter norms for participation in community-level activities and collaborative research, sharing of resources (such as stocks and tools) and distribution of data.

These and other features of model organism based communities do not follow from the epistemic roles model organisms play. But one can readily see that the epistemic features we have discussed both raise the need for, and support, a strong research community. The development and deployment of tools and methods is aided by collaboration. The ability to generalize from a result obtained in a model organism depends, in the ways we have discussed, on background information, which is easier to obtain when there is a supporting community structure, and when the outputs of its use can readily feed into future research. Thus, while we do not take a specific stand on how or to what extent model organisms anchor research communities, we believe that our discussion of the epistemology of model organisms can help explain the need for and the fruits of such communities, when they exist.

7. Conclusion

We have argued that model organisms are not theoretical models. While in a very broad sense both involve the use of tractable stand ins, these two strategies have distinct justificatory structures. Theoretical modeling is characteristically analogical: the modeler moves from a model to target on the basis of explicit and known features that they share. Model organisms, by contrast, are used as representatives of a broader class, justified either via "circumstantial" evidence or via phylogeny. There is also a difference in scope: while theoretical modeling occurs all over science, model organism inferences, especially when they are grounded in phylogeny, are restricted to biology.

Some may accept these arguments, but prefer to group theoretical models and model organisms under a common heading. As we discussed in section 6, model organisms play varied epistemic roles. It is not obvious which criteria to employ when deciding how to draw category boundaries in this kind of context. To our minds, it is a mistake to lump empirical extrapolation, such as those performed by Morgan and other biologists drawing on model organisms, with theoretical ones, like those performed by Volterra. We also think it helpful to set apart, where appropriate, the epistemic strategies characteristic of different parts of science. For these reasons, we favor a rather sharp demarcation between model organisms and theoretical models

Arnon Levy, Van Leer Jerusalem Institute, arnonl@gmail.com

Adrian Currie, Australian National University, adrian.currie@anu.edu.au

References

--- & Leonelli, S., [2011]: ‘What’s so Special about Model Organisms’, *Studies in the Histtory and Philosophy of Science*. 41: 313–23.

Ayala, Francisco J. [2009]: ‘Molecular Evolution vis-à-vis Paleontology.’ In *The paleobiological revolution : essays on the growth of modern paleontology*. D. Sepkoski and M. Ruse. Chicago: University of Chicago Press.

Bird, Alexander, [2011]: ‘Thomas Kuhn’, *The Stanford Encyclopedia of Philosophy*. E. N. Zalta (ed.), URL = http://plato.stanford.edu/entries/thomas-kuhn/#3.

Blattner, F.R. et al [1997]: ‘The Complete Genome Sequence of *Escherichia coli* K-12’, *Science* 277: 1453-1462.

Bolker, J.A. [1995]: ‘Model Systems in Developmental Biology.’ *Bioessays* (17)**5**, pp 451-455.

Bolker J.A. and Raff, R.A., [1997]:, 'Beyond worms, flies, and mice: It's time to widen the scope of developmental biology', *Journal of NIH Research*, 9: 35-39.

Burian, R. [1993]: How the Choice of Experimental Organism Matters, *Journal of the History of Biology*

Brigandt, I. and P. E. Griffiths [2007]: ‘The importance of homology for biology and philosophy.’ *Biology & Philosophy* 22(5): 633-641.

Carpenter, S. R. [1996]: ‘Microcosm experiments have limited relevance for community and ecosystem ecology.’ *Ecology* 77(3): 677-680.

Cleland, C. E. [2011]: ‘Prediction and Explanation in Historical Natural Science.’ *The British Journal for the Philosophy of Science*.

Crone, E., and J. Molofsky [1999]: Message in a Bottle? Utility and Limitations of Ecological Bottle Experiments, *Integrative Biology* 1: 209–214.

Currie, A. [2012]: Convergence as Evidence, *British Journal for the Philosophy of Science*.

Franklin, L.R. [2007]: Bacteria, Sex and Systematics, *Philosophy of Science*, 74: 69-95.

Frigg R. and Hartmann S. [2006]: 'Models in Science', *The Stanford Encyclopedia of Philosophy*. E. N. Zalta (ed.) URL: http://plato.stanford.edu/entries/models-science

Godfrey-Smith, P. [2006]: The Strategy of Model Based Science, *Biology & Philosophy* 21:725–740.

Gray, R. D. and Atkinson, Q. D. [2003]: 'Language-tree divergence times support the Anatolian theory of Indo-European origin.' *Nature* 426(6965): 435-439.

Gray, R. D, Greenhill, S.J. and Ross, R.M.. [2007]: 'The Pleasures and Perils of Darwinizing Culture (with Phylogenies).' *Biological Theory* 2(4): 360-375.

Griffiths, A.J.P., Wessler, S.R., Lewontin, R.C. and Caroll, S.B. [2008]: *Introduction to Genetic Analysis*, 9th edition, New York: W.H. Freeman.

Griffiths, P. E. [1996]: 'The historical turn in the study of adaptation.' *British Journal for the Philosophy of Science* 47(4): 511-532.

Hall, D. [2003]: 'Descent with modification: the unity underlying homology and homoplasy as seen through an analysis of development and evolution' *Biological Review* 78: 409-433

Hodgkin A.L. and Huxley, A.F. [1952]: ‘A Quantitative Description of Membrane Current and its Application to Conduction and Excitation in Nerve’, *Journal of Physiology*, 117 (4): 500-544.

Keller, E.F. [1996]: Drosophila Embryos as Transitional Objects: The Work of Donald Poulson and Christiane Nusslein-Volhard. *Historical Studies in Physics and Biology*, 26 (2): 313-346.

--- [2002]: *Making Sense of Life*, Cambridge, MA: Harvard University Press.

Kingsland, S. [1985]: *Modeling Nature: Episodes in the History of Population Ecology*, Chicago: University of Chicago Press.

Kohler, R. E. [1994]: *Lords of the Fly*, Chicago, IL: Chicago University Press.

LaFolette, H. and Shanks, N. [1996]: *Brute Science: Dilemmas of Animal Experimentation*. New York: Routledge

Leonelli, S. [2008]: Performing Abstraction: Two Ways of Modeling Arabidopsis Thaliana, *Biology & Philosophy*,23(4): 509-528.

Levy, A. [forthcoming], Modeling without Models.

Odenbough, J. [2006]: Message in the Bottle: The Constraints of Experimentation on Model Building. *Philosophy of Science* 73 (5): 720-729

Ma, Kevin Y., Chirrarattananon, Pakpong, Fuller, Sawyer B. and Robert J. Wood [2013]: Controlled Flight of a Biologically Inspired, Insect-Scale Robot, *Science*, 340: 603-607.

Morgan, T.H., Sturtevant, A.H., Muller, H.J. and C.B. Bridges, [1915]: *The Mechanisms of Mendelian Heredity*, London: Constable and Company.

Ratcliff, WC, Denison RF, Borrello M. and Travisano M. [2012]: Experimental evolution of multicellularity. *PNAS* Early Edition.

Rader, K. [2004]: *Making Mice: Standardizing Animals for American Biomedical Research*, 1900-1955, NJ: Princeton University Press.

Ramsey, G. and Peterson A.S., [2012]: 'Sameness in Biology.' *Philosophy of Science* 79(2): 255-275.

Reichenbach, H. [1956]: The direction of time. Berkeley: University of California Press.

Rosa, R. and Seibel, B.A.[2010]: 'Slow pace of life of the Antarctic colossal squid.' *Journal of the Marine Biological Association of the United Kingdom* 90(7): 1375-1378.

Santini, F & Stellwag, J. [2002]: 'Phylogeny, fossils, and model systems in the study of evolutionary developmental biology'. *Molecular Phylogenetics and Evolution*, Volume 24, Issue 3, Pages 379-383

Sansom, R. [2003]: 'Constraining the adaptationism debate.' *Biology & Philosophy* 18(4): 493-512.

Sober, E. [1988a]: *Reconstructing the Past: Parsimony, Evolution and Inference*, Cambridge, MA: MIT Press.

Sober, E. [1988b]: The Principle of the Common Cause. *Probability and Causality: Essays in Honor of Wesley C. Salmon*. W. C. Salmon and J. H. Fetzer. Dordrecht: Reidel.

Steel, D.P. [2008]: *Across the boundaries: extrapolation in biology and social science*. New York: Oxford University Press

Thomson-Jones, M. [2010]: Missing Systems and the Face Value Practice, *Synthese* 172 (2): 283-299.

Tucker, A. [2011]: Historical Science, Over- and Underdetermined: A Study of Darwin's Inference of Origins. *British Journal for the Philosophy of Science*. 62 (4): 805-829.

Volterra, V. [1926]: Fluctuations in the Abundance of a Species Considered Mathematically, *Nature* 118: 558-560.

Wade, M. J. [1977]: An experimental study of group selection. *Evolution* **31**: 134-135

Weber, M. [2005]: *Philosophy of Experimental Biology*, Cambridge: Cambridge University Press.

--- [2007]: Redesigning the Fruit Fly: The Molecularization of Drosophila. In: A. N. H. Creager, E. Lunbeck, and M. N. Wise (eds.), *Science without Laws: Model Systems, Cases, Exemplary Narratives*. Durham: Duke University Press, 23-45

Weisberg, M. [2007]: 'Who is a Modeler?' *The British Journal for the Philosophy Science*, 58: 207-233.

--- [2013]: *Simulation and Similarity: Using Models to Understand the World*, New York: Oxford University Press.

Wimsatt, W. [2000]: Generative Entrenchment and the Developmental Systems Approach to Evolutionary Processes. In: *Cycles of Contingency*, S. Oyama, P. Griffiths and R. Grey (eds.), Cambridge, MA: MIT Press.

Wolpert, L. [1969]: Positional Information and The Spatial Pattern of Cellular Differentiation, *Journal of Theoretical Biology*, 25 (1): 1-47.

[i] Volterra published his results in *Nature* in 1926. Unbeknownst to him, the American chemist and demographer Alfred Lotka had come up with essentially the same model in 1925 – hence the current label. We focus here on Volterra's work, because Lotka's was part of a grander project, attempting to systematize much of ecology and population biology (Kingsland, [1985]). Volterra's goal, at least initially, was more circumscribed, and, we think, more clearly an episode of model-based science. (Kingsland, [1985])

[ii] The ontological status of models is a complex question (Levy, [forthcoming], Thomson-Jones, [2010]). As noted earlier, our arguments do not depend on assumptions regarding ontology. At some points we speak of models as if they were objects of sorts – constructed and compared with empirical systems – but that is no more than a *façon de parler*.

[iii] Some philosophers use 'extrapolation' broadly to capture all manner of empirical inferences. We use the term more narrowly. In statistics and related areas, 'extrapolation' (and the closely related notion of 'interpolation') typically refers to an estimate of the relationship between two variables that moves from an interval in which their values have been observed to a larger interval in which they haven't. Our usage is close to this statistical notion, although we do not restrict it to the estimation of quantitative traits by specific statistical methods. On this usage, enumerative induction – in which one uses a sample in order to infer how common a certain type is within a given class ("How many Fs are Gs?") isn't a form of extrapolation. We thank Marcel Weber for a discussion of this point.

[iv] Morgan initially sought to study large-scale mutations, predicted by de Vries's theory, which were presumed to be very rare. Hence large numbers were of special importance.

[v] As a referee pointed out to us, the connection between cytology and Morgan's work is important: much *Drosophila* work might be extrapolable to 'chromosomal' organisms, rather than just 'sexual' organisms.

[vi] Steel ([2008]) calls this 'comparative process tracing'. His book contains a rich discussion of extrapolation in model organisms, but focuses on responses to some of the *objections* to such extrapolations, while we are targeting the relationship between model organisms and theoretical models.

[vii] They go on to note that the specific parameters might differ across different organisms.

[viii] Wimsatt ([2000]) calls this 'generative entrenchment'.

[ix] A discussion of epistemic constraints on phylogenetic inference would require a paper in and of itself. Some philosophers sceptical of model organism work have, by our lights,

underestimated the role phylogeny can play here (see for instance LaFolettee & Shanks [1996]). Other theorists are more directly concerned with phylogenetic reconstruction (Bolker [1995], Santini & Stellwag [2002]). We put aside a general discussion of such issues (but see Steel, [2008]).

[x] How to best conceptualize homology is contentious issue (Brigandt & Griffiths [2007]; Ramesey & Peterson [2012]; Hall [2003]). Here we draw on a 'taxic' account, which takes homology as an ancestral concept reliant upon systematics for epistemic access. This is largely for ease of exposition: the main points will fit with most other approaches.

[xi] Some would argue that HGT undermines the very concept of well-defined lineages in the prokaryotic world (Franklin, [2007]).

[xii] They are, perhaps, nearly unique to biology proper: the application of phylogenetic techniques to languages (Gray & Atkinson [2003]) and cultural products generally (Gray, Greenhill & Ross [2007]) suggest that something similar is at least possible in those domains.

[xiii] Often results of theory-guided experimental work at one point in time, serve to facilitate research at a later point in time – so they are "retrospectively" preparative.

[xiv] Kuhn uses the notion of 'paradigm' in several senses. We only employ one, and do so in a qualified way. (Bird, [2011] §3).

[xv] It should be noted that some have argued that features particular to yeast biology have affected the result obtained by Ratcliff et al – but as one would expect (on the basis of our discussion) this is taken as telling against the theoretical significance of their results.

[xvi] We thank an anonymous referee for raising this point.

Captions

Figure 1: Phylogenetic Inference. In the first step a character is inferred from some sample lineage to an ancestor of that lineage. In the second step we infer from that ancestor to a target lineage, or the clade in general.

Acknowledgements

For written and oral comments we are indebted to: Rachael Brown, Peter Godfrey-Smith, Paul Griffiths, Sabina Leonelli, John Matthewson, Emily Parke, Kim

Sterelny, Marcel Weber, David Wiens, an audience at the 2012 NZAAP in Wellington and an anonymous referee for this journal.