sciendo

Are Model Organisms
Theoretical Models?

Veli-Pekka Parkkinen
University of Bergen

Abstract
This article compares the epistemic roles of theoretical models and model organisms in science, and specifically the role of non-human animal models in biomedicine. Much of the previous literature on this topic shares an assumption that animal models and theoretical models have a broadly similar epistemic role—that of indirect representation of a target through the study of a surrogate system. Recently, Levy and Currie (2015) have argued that model organism research and theoretical modelling differ in the justification of model-to-target inferences, such that a unified account based on the widely accepted idea of modelling as indirect representation does not similarly apply to both. I defend a similar conclusion, but argue that the distinction between animal models and theoretical models does not always track a difference in the justification of model-to-target inferences. Case studies of the use of animal models in biomedicine are presented to illustrate this. However, Levy and Currie's point can be argued for in a different way. I argue for the following distinction. Model organisms (and other concrete models) function as surrogate sources of evidence, from which results are transferred to their targets by empirical extrapolation. By contrast, theoretical modelling does not involve such an inductive step. Rather, theoretical models are used for drawing conclusions from what is already known or assumed about the target system. Codifying assumptions about the causal structure of the target in external representational media (e.g. equations, graphs) allows one to apply explicit inferential rules to reach conclusions that could not be reached with unaided cognition alone (cf. Kuorikoski and Ylikoski 2015).

Keywords
Biomedicine, modelling, animal models, theoretical models, scientific representation.

1 Introduction

How does studying a model organism create understanding of its target? Can both model organism research and theoretical modelling be understood as a type of indirect representation; the study of a surrogate system from which results are transferred to the target based on known analogies between the two systems? Recently, Levy and Currie (2015) have argued to the contrary, concluding that animal models differ from theoretical models in the crucial point of justifying model-to-target inferences. According to Levy and Currie, instead of serving as a surrogate for its target in virtue of a deliberately constructed analogy, a model organism serves as a basis for empirical extrapolations that are justified by circumstantial empirical evidence or phylogenetic inferences.

This article considers the epistemic role of model organisms in light of Levy and Currie's taxonomy of models. I present examples of the use of non-human animal models in biomedicine that illustrate how scientists can and do employ reasoning strategies analogous to those that Levy and Currie associate with theoretical modelling. These strategies include deliberate construction of analogies between the model and the target, and the evaluation of the robustness of results derived from a family of models. I intend this not as an argument against the basic point made by Levy and Currie, however. Instead I argue that theoretical models do have epistemic characteristics that differ from those of animal models, as Levy and Currie suggest, but this distinction does not robustly track differences in the strategies of justifying model-to-target inferences. I argue that the distinguishing feature of theoretical models vis-à-vis animal models and other concrete model systems is their affordance to be used as inferential aids. Theoretical modelling typically involves the use of some external representational media—graphs, equations and the like—to facilitate inferences about a target system based on well specified assumptions that are explicitly encoded in the model. The results obtained through the study of a theoretical model are true insofar as the assumptions coded into the model are true of the target. By contrast, the causal mechanisms responsible for the properties and behaviour of an animal model are (obviously) not explicitly coded in the model, but are rather something that one endeavours

to discover by studying the model, and then extrapolate to the target. Animal models thus serve as surrogate systems in a more literal sense: they provide empirical evidence for hypotheses about the properties and behaviour of their targets based on an assumption of causal similarity between the two systems.

The structure of the article is as follows. Section 2 introduces the prevalent philosophical conception of modelling which sees model-based research as a form of indirect representation through the study of a surrogate system, wherein the model-target relationship is grounded in deliberately manufactured analogies between the two systems. Section 3 reviews recent philosophical work on model organisms. Section 4 presents Levy and Currie's arguments against subsuming model organism research under the surrogate system view of theoretical modelling. Section 5 presents case studies from model organism work in biomedicine to illustrate how methods that Levy and Currie' associate with theoretical modelling may also be used in research on animal models. Section 6 defends Levy and Currie's thesis about the difference between theoretical modelling and model organism research through a different route, arguing that theoretical models play a specific role as inferential tools whereas animal models serve the role of a surrogate source of evidence. Section 7 summarizes the conclusions.

2 Models as surrogate systems

The focus of philosophers studying model-based reasoning has mostly been on theoretical modelling. Some accounts identify theoretical models with set theoretic structures or other abstract objects, while others analyse models into narratives or fictional objects that, if they were real, would be concrete systems in the world (Suppe 1989: ch. 1, Giere 2004, Frigg 2006, Godfrey-Smith 2009, Morrison 2009). Such analyses differ in terms of how they cash out the representational properties of models, the role of false assumptions and idealization in modelling, the edacity of model-world-modeller relations, and explanatory power of models. Despite the differences, virtually all accounts are compatible with a picture of the practice of modelling that can be called the 'surrogate system' view, which sees modelling as indirect representation of a part of the world through

the study of an artificial system that is used as a stand-in for its target.

In this view, modelling consists of analysis of a surrogate system that is simpler and easier to manipulate than its intended target, for the purpose of discovering results that one could not feasibly produce by studying the target system directly. Model building starts with an initial idea about the causal structure of the target system, the crucial parts of which the modeller intends to replicate in the model. The behaviour of the model is then studied under various manipulations of its parts, and compared to the observed behaviour of the target, leading to subsequent refinements of the model. Once a stage where a satisfactory resemblance between the model and the target has been achieved, the model takes a role roughly analogous to an experimental system studied in a laboratory: the modeller produces results in an artificial modelling environment that can be directly controlled, and applies these to the target system 'in the wild'. Mäki (2005) argues that modelling is not merely analogous to experimentation, but models and experiments provide literally the same kind of epistemic output—both provide piecemeal insights into aspects of the target system by isolating specific causal mechanisms from interfering factors that would make studying the mechanism difficult in its natural context. The difference between models and experiments is in how this 'shielding' of the mechanism of interest from interfering factors is achieved: either by abstracting away from possible interfering details in a theoretical model, or through careful control of the causal setup of an experiment.

The central question for the surrogate system view is how the transfer of results from the model to the target is justified. While experiments are performed in conditions that retain some of the material features of the target system and can thus be taken to provide valid results concerning the operation of just those parts of the target that are replicated in the experiment, theoretical models cannot be validated in this way. To explain how theoretical models apply to their targets, the surrogate system view takes the description of a theoretical model (text, equations, diagrams, etc.) to specify either an abstract object, or a concrete but fictional object that stands in some type of similarity relation to its target. These similarities ground an analogy between the model and the target in some core features of interest, on which the representational capacity of the

model rests.

For example, the model specified by the Lotka-Volterra equations is taken to replicate the influence of predation on predator and prey population growth rates as a mathematical dependency between the model's variables. The more closely this mathematical dependency maps to the dependencies that govern the abundance of actual predator and prey populations, the greater the validity of the model. To improve the analogy between the model and the target, the modeller may add more components to the core model to make it capture more causally relevant facts about its target, such as adding terms that stand for the carrying capacity of the ecology or for the effects of intraspecific competition on the growth of the populations.

In addition, justifying the step from the model to the target requires a way to control for the distorting influence of those features of the model that are deliberately idealized or cannot be validated due to lack of empirical evidence form the target. The frequent use of robustness analysis by modellers can be seen as playing this role (Weisberg 2004). Robustness analysis consists of the construction of a family of models that vary in other parts than the core features purported to represent the operation of the mechanism of interest, and searching for results that concord across the whole family of models. This allows the modeller to isolate results that are independent of unrealistic modelling assumptions from those that are idiosyncratic to particular model specifications.

Although an operation performed exclusively on models, when combined with suitable empirical evidence to validate the core structure of the model, robustness analysis can yield model-based confirmation of hypotheses about the target by controlling for the distorting effect of idealizations and omissions. Consider an example: The cyclical behaviour of predator and prey populations that the Lotka-Volterra model exhibits might be an idiosyncratic feature of the model, entirely dependent on the specification of the functional form of the equations chosen by the modeller. If this were the case, results derived from the model would not represent the behaviour of any target ecology, as they would merely reflect a fact about the particular mathematical dependency. But it can be shown that certain behaviours of the Lotka-Volterra model are stable against variations in the functional forms of its equations. Combined with evidence

confirming the relatively modest modelling assumption that preda-
tor abundance depends on prey abundance which crucially depends
on predator death rate, one can be reasonably confident that the
model reproduces real consequences of such predator-prey interac-
tions despite the analogy between the model and the target being
incomplete (Weisberg 2004).

We can now summarize the main strategies of justifying model-
to-target inferences in theoretical modelling, as understood by the
surrogate system view. The application of a modelling result to a
real world target rests on an analogy between the model and target
systems, established through explicit comparisons of their features.
The degree of analogy is entirely controlled the modeller, who may
add or remove features from the model depending on her research
interests. However, in order for the model to be tractable, some
features of the target are deliberately omitted or simplified in the
model. To control for errors that are due to idealization, omission,
and the underdetermination of modelling assumptions by empirical
data, one may employ robustness analysis to identify results that are
independent of problematic assumptions.

The surrogate system view makes no in-principle distinction be-
tween abstract mathematical or computational models and concrete
physical models such as scale models used in engineering. Like ab-
stract models, concrete model systems serve as surrogates for their
targets, being constructed to facilitate manipulations that could not
be done in the target system directly. Just as in theoretical model-
ling, the validity of a result obtained by studying a concrete model
depends on how faithfully the model system can mimic what would
happen if corresponding manipulations were performed directly on
the target system. This analysis can be effortlessly extended to model
organism research—a model organism is an experimental system
that stands in for its target in virtue of a partial analogy between the
model and the target. In the literature that touches upon or explicitly
focuses on model organism research, this move is taken by promi-
nent authors such as Weisberg, and Ankeny and Leonelli, the lat-
ter relying on Morgan and Morrison's framework in their analysis of
model organism research (Ankeny 2009, Ankeny and Leonelli 2011,
Morgan and Morrison 1999, Weisberg 2013).

The surrogate system view thus promises to provide a unified

analysis of diverse types of model-based reasoning, from theoretical models to concrete model systems and model organism research. In each case, a model is a surrogate representation of a target which replaces direct experimentation on the target itself, and the inferences from the model to the target are grounded in the deliberately constructed analogy between the model and the target.

3 Previous philosophical work on model organism research

Much of the previous philosophical work on the use of laboratory organisms in science rests on some version of the idea that models are surrogate systems that represent a target 'in the wild', i.e. that the model organism is a system that the researcher interprets as standing for the target organism(s) of interest. Thus conceived, model organisms could be subsumed under the view that takes both theoretical modelling and material models to be forms of indirect representation through the construction and study of surrogate stand-ins for the real system of interest. For example, Ankeny and Leonelli argue that all experimental organisms can be understood as models qua the 'models-as-mediators' framework of Morgan and Morrison. For Ankeny and Leonelli, proper model organisms are a subset of such experimental organisms (Ankeny and Leonelli 2011: 315–6).

According to Ankeny and Leonelli, the subset of experimental organisms that are model organisms is individuated by reference to the intentions of the researchers studying those organisms, and by the social attributes of the research community that studies them, such as practices of sharing data and methods:

> [model organisms] are explicitly models for whole organisms, to be investigated using a range of disciplinary approaches with the intention of integrating these approaches to develop an understanding of the whole, intact organism. Hence the epistemic goal shared by members of the model organism communities clearly has shaped the way in which model organisms are not only viewed, but also used as models. (Ankeny and Leonelli 2011: 320)

> [...] model organism research programs share a specific interpretation of the extent of the representational scope of their model organisms. [...]In sum, model organisms are models for whole organisms, whose

potential representational scope extends to all living beings. (Ankeny and Leonelli 2011: 320)

Ankeny and Leonelli thus define 'model organism' as a status concept that is attributed to experimental organisms within a disciplinary context characterized by particular explanatory goals and institutions that support the integration of various disciplinary approaches. The distinction between experimental organisms and proper model organisms is interesting and important. However, for my purposes, the most important point is that by drawing on Morgan and Morrison's analysis, Ankeny and Leonelli conceive all experimental organisms as akin to theoretical models in a certain very general sense—experimental organisms are surrogate systems interpreted as stand-ins for phenomena outside the laboratory, much like theoretical models are abstract stand-ins for real world phenomena. The set of proper model organisms is individuated by the intentions of the researchers and the social organization of the relevant research community, not by differences in the justification of model-to-target inferences.

Jessica Bolker (2009) argues for a slightly different taxonomy of experimental organisms. Bolker draws a distinction between experimental organisms that serve as exemplars of a kind, called exemplar models, and organisms that serve as surrogates for a specific target, called surrogate models. According to Bolker, exemplar models are mainly used for learning about fundamental biological processes generalizable to a large group of organisms. Thus, exemplar models are typically chosen based on phylogenetic reasoning—the greater the degree of conservation between the model and its target(s), the greater the validity of the model. According to Bolker, this use of model organisms is characteristic of basic research, which aims to understand fundamental biological processes found in a wide range of organisms. By contrast, surrogate models are used for creating understanding of a specific target. The most obvious example is the use of experimental animals in biomedicine, where the purpose of studying an animal model is to learn specific clinically relevant causal facts about humans. According to Bolker, what matters for the choice of a good surrogate model is the de facto similarity of the model and its target in the particular trait of interest, not the generalizability of

the model results across taxa. Thus, in contrast to exemplar model research, it does not matter whether the trait of interest is conserved across taxa or not, as long as the model, for whatever reason, exhibits traits comparable to those of interest in the target.

How does Bolker's analysis relate to the ideas introduced in the previous section about model-based reasoning as indirect representation through surrogate systems? Bolker occasionally talks of both exemplar and surrogate models as representations of their targets. However, it is clear that she intends this to be understood differently for the different types of model organism. In Bolker's taxonomy, exemplar models are representatives of a class of organisms: results obtained in exemplar models are evidence for generalizations about the species or higher taxa to which they belong. That is, the inference from an exemplar model to the target is induction from a sample. Exemplars are thus 'models' only in a very permissive sense of the term—one in which instances of a kind 'model' the kind by supporting induction of the model's properties to other instances of the same kind. By contrast, surrogate models *substitute for* their intended target—a surrogate model is interpreted as a stand-in for its target qua directly established similarities, and studied as if one were studying the target itself (Bolker 2009: 486, 489, 495). This latter notion of modelling is much closer to the surrogate system view common from the literature on theoretical modelling.

Huber and Keuck offer an analysis of animal models in biomedicine that is reminiscent of Bolker's account of surrogate models (Huber and Keuck 2013). According to Huber and Keuck, model organisms in basic biological research are intended as evidence for general conclusions about general biology (Huber and Keuck 2013: 386). This role is contrasted with the role that animal models play in biomedicine, where models are said to substitute humans as research subjects for studying clinically relevant phenomena in humans (Huber and Keuck 2013: 386). While the former are conceived of as basis for generalizations much like Bolker's exemplar models, the epistemic function of the latter class of animal models is analysed using a much stronger notion of representation: According to Huber and Heuck, animal models in biomedicine are (material) models of abstract descriptions or hypotheses of disease-processes (Huber and Keuck 2013: 386). This is much in accordance with the models-as-mediators account of

Morgan and Morrison and thus with the basic framework of modelling as indirect representation. Huber and Keuck add that, due to their nature as living beings, animal models in biomedicine could be conceived as analogue models of human disease processes (Huber and Keuck 2013: 389).

The above exposition admittedly fails to represent the full breadth of philosophy of science literature on model organisms, and glosses over many nuances of the views presented. The purpose of this brief sampling of views is merely to highlight the fact that prominent analyses of the epistemic role of model organisms share a view that either all experimental organisms (for Ankeny and Leonelli), or at least an important subset of them in biomedicine (for Bolker, and Huber and Keuck) function akin to theoretical models in the sense that inferences from the model to the target are based on some sort of robust representation relation. In the next section, I move on to discuss at some length a view that disagrees on this point, and argues that model organisms and theoretical models *differ* in an important way in the justification of model-to-target inferences.

Before moving on, it should be noted that 'model organism' clearly means slightly different things for different authors. For the purposes of this paper, I will not adopt a stance on what is the correct usage of this term in relation to more general notions such as 'experimental organism'. Fine grained distinctions between model organisms and other living experimental systems, as important as they are, are not crucial for the comparison between theoretical models and laboratory organisms with respect to the justification of model-to-target inferences, which is the main focus of this paper. In what follows I will discuss the reasoning strategies employed in biomedical use of laboratory animals in the light of Levy and Currie's taxonomy of models, and will therefore use the term 'animal model' to refer broadly to any laboratory animal studied for the purpose of making clinically relevant inferences about humans.

4 Levy and Currie on model organisms and theoretical models

In a recent article, Arnon Levy and Adrian Currie compare the epistemic characteristics of paradigmatic theoretical models and model organisms in biology, and argue that it is a mistake to subsume both

under a unified account of modelling as indirect representation (Levy and Currie 2015). Levy and Currie accept the surrogate system view as an account of theoretical modelling, but argue that model organism research differs importantly from theoretical modelling in the justification of model-to-target inferences. Levy and Currie's target is the idea of explicit feature mapping that is required for establishing an analogy between a (theoretical) model and its target. In theoretical modelling, this is feasible precisely because the properties of the model are entirely controlled by the modeller, who may improve the analogy in crucial parts of the model and loosen it in less important parts, and may freely interpret parts of a model as standing for parts of different target systems according to the aims of the research. By contrast, a scientist working with an animal model often has no comparable control over the degree of analogy between the model and the target, and may not know enough of the target to improve the analogy even if such control were possible—the purpose of studying a model organism is often to establish that the target has some properties that we cannot directly measure.

Levy and Currie argue that the justification of inferences based on model organism research rests on supporting empirical arguments rather than an analogy established by direct comparison of the model and the target. These supporting arguments are either based on circumstantial empirical evidence, or on phylogenetic evidence that establishes relevant homologies between the two organisms. In the first case, evidence from a range of target organisms is available, demonstrating coarse similarities in the feature of interest, lending simple inductive support for inferring the same feature in the target. Note that Levy and Currie's point here is about the justification of the inference from the model to the target, not about how scientists actually choose models. Many kinds of circumstantial considerations other than phylogeny often play a role in model choice, many of which have to do with the experimental accessibility and other practical features of the model system, as well as features strictly speaking outside the model, such as standardizing the environment in which the model is studied (see e.g. Ankeny et al. 2014 for an extended case study).[1]

---

[1] I owe this point to an anonymous reviewer.

The phylogenetic strategy is a bit more complicated, involving two inferential steps supported by empirical evidence. First, observing similarities in a trait between the model species and closely related lineages other than the target species, one infers that a common ancestor exhibited the same or similar trait that is found in the model, as traits that are found in many lineages in a clade are likely to have evolved from the same ancestral trait. The trait is then projected to the target in which it cannot be directly observed, based on the knowledge that the target species shares an ancestor with the model species.

In the case of deeply entrenched, evolutionarily basic traits, phylogenetic arguments can be extended further across species boundaries (Weber 2005). This enables biologists to theorize about a wide range of organisms based on a limited stock of well-understood models, while the strength of the model-to-target extrapolation depends on two factors: the entrenchment of the trait, and the distance between the model and the target in the phylogenetic tree. The key difference to theoretical modelling is thus that experimental organisms earn their status as surrogate systems through additional empirical arguments, not through explicit feature mapping that establishes analogies between the model and the target. The model organism is also typically intended as a type-specimen of a wider, less sharply bounded range of species, rather than being chosen with a particular target system in mind. Levy and Currie's analysis thus comes close to what Jessica Bolker calls exemplar models—models chosen for their ability to support generalizations to a wide range of organisms due to conservation of traits. However, Levy and Currie clarify the point that even though the model-to-target inferences are ampliative, they are not straightforward inductive inferences as Bolker describes them. The first step in a phylogenetic argument is an abduction: a common ancestral form of a trait is inferred from the fact that common descent would be the best explanation of trait similarities in a clade. Once common ancestry is abductively established, the trait of interest is then projected to the target organism of interest.

Levy and Currie highlight an important strategy of model-based reasoning in biology whose epistemic characteristics cannot be accounted for by extension of the surrogate system view of theoretical modelling, and thus conclude that model organisms in biology are

not (like) theoretical models. For Levy and Currie, model organisms are not used to represent anything, but function as a source of empirical evidence that, together with other supporting evidence and arguments, allows extrapolation to the target. However, animal models are also widely used outside basic research in biology, and it is not clear that the difference in strategies of justifying model-based inferences applies the same way across diverse practices of animal model research. Scientists working with animal models may quite flexibly employ any of the strategies described above—deliberate control of analogy and robustness reasoning, circumstantial evidence, or phylogenetic arguments—or some combination of them, depending on the available models and the questions one wants to answer using the model.

In particular, examples of such mixed practices can be found in biomedicine, where common animal models are studied with a specific target (humans) in mind, the models are often not closely related to their intended target, and the research questions may consider high level physiological or even psychological mechanisms that are not evolutionarily basic traits. This makes it difficult to validate the animal models by any straightforward application of phylogenetic reasoning or through a search for circumstantial evidence across many species. To illustrate the point, the following section describes two cases where strategies that in Levy and Currie's taxonomy would associate with theoretical modelling are applied in research on animal models. These case studies are not intended as a historical argument against Levy and Currie. Rather the purpose is merely to point out that it is feasible to use a mixed package of strategies to justify inferences from an animal model to a specific target: the methods by which an experimental organism's surrogate status is justified do not distinguish animal models in general from theoretical models.

5 Control of analogy and robustness reasoning in model organism research

Levy and Currie point out an important distinction between animal models and theoretical models, having to do with the justification of model-to-target inferences. However, as this section illustrates, the

difference in justification strategies pointed out by Levy and Currie does not cut neatly across theoretical models and model organisms. While many experimental organisms gain their status as models through considerations of homology or circumstantial empirical evidence, these are not the only strategies used. Purposive control of model-target analogies and robustness reasoning play a role also in model organism research, as they do in the practice of theoretical modelling.

5.1 Direct control of (dis-)analogy

Maugeri and Blasimme (2011) employ examples from cancer research to argue that research employing animal models in molecular medicine relies on extensive engineering of models in order to create positive analogies and remove negative ones between the models and their human targets. Their case studies are murine models of human cancer, where extensive preparation and genetic manipulation is used to eliminate sources of disanalogy between the animal model and humans. Maugeri and Blasimme discuss two ways this is done in mouse models of cancer. One has to do with the preparation of so called xenograft transplantation models: immunosuppressed mice made to host human cell grafts. The other involves genetic manipulation of mice to induce tumour growth through pathways homologous to human counterparts.

In the first case, the native immune system of the mouse is depleted to prevent the rejection of cell transplants from humans. This permits the transplantation of human stem cells into the recipient mouse, where the mechanisms implicated in tumorigenesis can be studied in vivo, enabling experiments that would not be possible in isolated cell lines. In this case, a potential source of disanalogy comes from the fact that the murine immune system is prevented from interacting with the transplanted human cells, yet this enables the researchers to establish novel results in an environment that more faithfully reflects the conditions in which actual human cancers develop (Maugeri and Blasimme 2011: 611–2). Note that one might argue that the hybrid humanised mouse model does not count as a model in the proper sense: what is analysed in the hybrid organism is the behaviour of parts grafted directly from the target system.

Here the model organism merely provides a host environment that allows experiments that would not be possible in isolated human cell cultures, and the purpose of engineering the model is to prevent the host physiology from contaminating the natural behaviour of the transplanted human parts. This however does not undermine the basic point: the purpose of engineering the experimental organism is to facilitate the extrapolation of results from one system to another by eliminating a potential source of disanalogy between the two systems.

Maugeri and Blasimme's other example is the genetically modified mouse model (GEMM). GEMMs became a tool for cancer research when it was shown that transgenic mice could be disposed to cancer by mutating specific oncogenes. GEMMs enable researchers to study tissue-specific cancers by triggering tumorigenesis through activating oncogenes and depleting tumour-suppressor genes in the tissue of interest. The tumours so induced are derived from the ingenious murine cells, and thus an assumption is needed that the mechanism of tumorigenesis is analogous between mice and humans. This assumption is validated partly by appeal to evolutionary ancestry, and partly by deliberate engineering of the model. The GEMM mice are used in the study of human cancer by targeting interventions to genes that are orthologous to human oncogenes and tumour-suppressor genes, and by deliberately generating lesions that mimic those found in human cancer cells. The GEMM mice are thus validated by a mixed strategy of appeal to evolutionary ancestry, and the creation of positive analogies to humans at the genetic level (Maugeri and Blasimme 2011: 612–3).

In the cases reviewed by Maugeri and Blasimme, the experimental animal is not merely a sample specimen used as basis of induction to a broader class of organisms, but a system intentionally engineered to mimic a specific causal structure of interest: cancer biology in humans. The models are engineered to be as faithfully analogous as possible to the specific target—xenograft transplantation models may even be used to study the cancer of a specific individual patient—rather than to maximize the inductive scope of the model. The models thus qualify as proper surrogate systems from which evidence is extrapolated to human targets. The working strategy of the researchers studying mouse models of cancer proceeds through

rounds of comparisons between the model and the human target, during which the model is deliberately engineered until a satisfactory resemblance is achieved that would allow drawing conclusions about humans based on the results seen in mice. This involves both modifying the mouse physiology to remove obstacles to extrapolation, as well as incorporating novel components from the target to the mouse model.

Similar practices have been described in other work on mouse models as well (Nelson 2013). In the process of preparing and refining a model, observations of the clinical phenomenon of interest in humans lead to modifications of the model, but the results obtained in the model also influence the criteria by which the target phenomenon is individuated in humans. Nicole Nelson has described such feedback between modifications of the model and adjustments of the clinical state description as epistemic 'scaffolding'; that of constructing the grounds for making successively more ampliative, and hence epistemically risky, claims (about humans) based on the results obtained in an animal model (Nelson 2013). This process resembles the process by which the specification of a theoretical model is adjusted through successive rounds of studying the model and comparing the results to the real world target. In comparison, the reasoning strategies outlined by Levy and Currie seem to play a smaller role in such scaffolding.

5.2 Robustness reasoning

Explicit comparison and modification of the model with respect to its target is not always an option for researchers working with animal models. When explicit feature mapping is not an option, scientists working with animal models may still appeal to assumed analogies indirectly, by attempting to discharge errors due to expected disanalogy by applying robustness reasoning. This strategy, too, is analogous to theoretical modelling. The role of robustness arguments is illustrated in the discovery of cholesterol's role in the mechanism of atherosclerosis. Atherosclerosis is a condition affecting coronary arteries, characterized by lesions that contain large numbers of macrophages and the proliferation of smooth muscle cells. Its clinical manifestations are infarction or stroke that occur when a hardened

fibrous plaque develops on top of the lesion occluding blood flow locally, or rupturing from the vessel wall and obstructing narrow vessels elsewhere. It is currently held that an elevated level of low density cholesterol carrying lipoproteins in the blood is a major causative factor in atherosclerosis.

The earliest evidence indicating that cholesterol may trigger the development of atherosclerosis came from animal experiments conducted by a Russian pathologist Nikolai Anitschkow. Anitschkow showed that that rabbits kept on a high-cholesterol diet developed lesions in the arterial wall similar to those found in human patients diagnosed with heart disease (Anitschkow 1913, 1933). At the time of these findings, atherosclerosis was not recognized as a distinct disease category. Instead, the thickening and hardening of arteries was considered an inevitable side-effect of aging and natural 'wear and tear'. Anitschkow's results suggested that contrary to the received view, atherosclerosis might be a preventable and perhaps even reversible condition resulting from prolonged high blood cholesterol level. Unsurprisingly, the results were received with considerable scepticism.

The scepticism was not unfounded. To transport a causal relation from an animal model to humans, one would have to assume that the model is analogous to humans in the mechanisms that mediate the cause and the effect. But the details of the mechanisms linking dietary or blood cholesterol to atherosclerosis were largely unknown at the time—the cholesterol theory of atherosclerosis was a completely novel hypothesis. Furthermore, there were independent theoretical reasons to believe that rabbits and humans were not causally analogous in the relevant sense. Rabbits are natural herbivores, not accustomed to lipid-rich diet the same way humans are. If cholesterol indeed had the capacity to trigger atherosclerosis, one would expect rabbits to be much more sensitive to its effects than humans are. Until further details of the mechanisms implicated in the development of atherosclerosis were articulated, the validity of the early experiments as evidence about human atherosclerosis was contestable.

This however did not prevent a number of experimentalists from picking up on Anitschkow's work. Several other animal models of cholesterol-induced atherosclerosis were developed in the following decades, and comparative chemical analyses of human and animal

atherosclerotic lesions were conducted based on these results (Aschoff 1933, Hirsch and Weinhouse 1943, Leary 1944, Bruger and Oppenheim 1951). The degree of analogy between the animal models and humans was somewhat an open question, due to the aforementioned lack of knowledge of the pathophysiology of atherosclerosis. What one could do to secure the analogy, however, was to compare the outcomes of cholesterol interventions across a range of animal models that are known to differ from each other in cardiovascular physiology and natural dietary habits; two of the most likely variables to modulate the effect of the intervention. If the results were stable across the variable causal background of many dissimilar models, this would show that cholesterol's atherogenic effect is not dependent of a mechanism that is specific to a particular experimental organism. This amounts to empirical robustness analysis: an attempt to isolate modelling results that are not affected by negative analogies by looking for concordant results in a family of models in which potential sources of disanalogy are expected to vary. Note that this is not the same as to establish a mechanistic analogy between the animal models and humans. Robustness analysis simply allows one to be more confident that specific (even unknown) disanalogies between the models and humans do not compromise extrapolation.[2]

The degree to which the animal models could robustly demonstrate the atherogenetic effects of cholesterol became an important point of contention in the debate over cholesterol's role in atherosclerosis. Louis Katz, a major figure in American cardiology and an early proponent of the cholesterol hypothesis summarized the output of the model organism studies as follows:

> Even the most severe critics of the cholesterol concept of atherosclerosis cannot fail to note the conspicuous similarities between cholesterol-induced lesions—whether in rabbit, dog or chick-and the lesions seen in man. Taking into consideration the anatomic differences among the various species in architecture of the great vessels, the similarities between experimental cholesterol-induced and human lesions are indeed remarkable. (Katz 1952: 104)

Katz inferred that the concordance of results from heterogeneous models with clinical observations of human patients sufficed to con-

---

[2] I thank an anonymous reviewer for clarifying this point to me.

firm a causal relation between cholesterol and atherosclerosis in humans (Katz 1952: 102).

Others contested the qualitative similarity between human atherosclerosis and the experimental models. As an example, an authoritative clinical chemistry textbook by John Peters and Donald Van Slyke points out that the distribution of the lesions in the animal models does not match what is seen in humans, and that the models differ greatly in their sensitivity to various factors that modulate the development of atherosclerosis in response to excess cholesterol (Peters and Van Slyke 1946: 534–5). If the positive results were indeed not independent of the species-specific physiological and anatomical features of the experimental animals, then the credibility of the positive findings as evidence of causality in humans would be contestable. Peters and Van Slyke consequently conclude that the evidence does not license more than a contributing role to cholesterol in the development of atherosclerosis. Similar observations were made by other authors: the cholesterol intervention would reliably produce atherosclerosis in herbivorous animals such as rabbits and hamsters, but results in other species such as rats and dogs that one could expect to better reflect human cholesterol metabolism and cardiovascular anatomy were much more variable (Duff and McMillan 1951).

These observations would undermine the robustness of the findings in animal models, calling into question the validity of the models. The status of the model studies indeed remained contestable until details of the mechanism became better understood, allowing researchers to compare intermediate endpoints in the disease process to establish relevant analogies. Model organism research nonetheless proved pivotal in establishing the now widely accepted view that cholesterol plays a causal role in the development of atherosclerosis, and various animal models including rabbits are still used in the field. In the early days, little was known of the causal mechanisms responsible for atherosclerosis. Instead of establishing analogies by explicit feature mapping, the validity of the models were assessed by the use of robustness reasoning in an attempt to discharge errors that are due to inevitable causal disanalogies between humans and the model organisms.

Both feature mapping and robustness reasoning aim to provide evidence for the reliability of the extrapolation from the concrete

model system to the target. For model-based conclusions about humans to be true, the mechanism in humans and in the model must not be so dissimilar as to introduce differences between the model and humans in the outcome of interest. Engineering the model to exhibit features similar to human equivalents can be seen as providing a reason to believe that the model-based conclusions are true by way of creating direct evidence of mechanistic similarities. But model-based conclusions can be true even if one has no reason to firmly believe so, i.e. if one does not know the relevant mechanistic similarities. All that matters is that the mechanisms are in fact similar enough. Robustness analysis works by ruling out explanations of model results in terms of mechanisms idiosyncratic to the model, thus supporting extrapolation without directly establishing that the mechanisms are similar. Both strategies seem to have analogues in theoretical modelling. Explicit feature mapping by refining the specification of the model is evidently used to validate theoretical models. Likewise, robustness analysis is used in theoretical modelling to rule out the possibility that modelling results crucially depend on background assumptions that cannot be directly verified.

It should be stressed that these cases are not, by far, representative of all animal model research in biomedicine.[3] My claim is not that the reasoning strategies exemplified by these cases exhaust the ways that biomedical scientists deal with the problem of transferring results from an animal model to humans. Rather the claim that these examples are brought to support is that the reasoning strategies that Levy and Currie associate with theoretical modelling *can* be used in the context of working with an animal model.

6 Models as sources of evidence, and as inferential aids

So far I have pointed out that controlling model-target analogies and applying robustness analysis are used to justify model-based inferences both in theoretical modelling and model organism research. This would undermine the basis of Levy and Currie's taxonomy of

[3] While obviously not an exhaustive sample of biomedical use of animal models, these research areas—atherosclerosis and cancer research—are undoubtedly major fields within biomedical research.

theoretical versus animal models, but not their claim about there being a difference in the epistemic output of theoretical modelling and animal model research. This section clarifies the point about the epistemic roles of these two types of models, drawing attention to the specific use of theoretical models as representational tools that facilitate inferences about their targets based on well specified modelling assumptions.

In the surrogate system view, model-based reasoning proceeds through comparisons between the model and target systems. Models are validated by establishing analogies, and new things are learned about the target by observing the behaviour of the model and then exporting the result to the target. It is thus not surprising that laboratory research on model organisms is seen as serving a fundamentally similar purpose as theoretical modelling—both practices involve some sort of (quasi-)experimentation and extrapolation of results from the model to the target. Levy and Currie accept the rudiments of this picture, but claim that the crucial difference is in the justification of the extrapolation; in model organism research this requires further empirical evidence in support of the phylogenetic argument. This is an important strategy, but the examples presented above hopefully demonstrate that the validity of results obtained in animal models can be justified with strategies analogous to those that are paradigmatically associated with theoretical modelling.

Despite this, there are still reasons to think that animal models differ from theoretical models in their epistemic profile. Levy and Currie stress that model organism research always involves some sort of extrapolative inference, transferring the evidence created in the model to bear on the target. It might appear that theoretical modelling involves a similar inductive step, the difference being that the inference is justified by the analogy that the modeller has verified beforehand, instead of relying on an assumption of causal analogy that must be justified by further empirical evidence. Although this is a heuristic that theoretical modellers appear to follow, it is not altogether clear that it explains the epistemic role of theoretical models in a satisfactory way. Consider how exactly could the study of an abstract or fictional entity create new empirical knowledge about the target, over and above the knowledge of the properties of the target that is used in constructing the model?

However this may be, there is an epistemic role that theoretical models undeniably play that does not hinge on the model's surrogate role. This role, which is distinct to theoretical models only, is to extend the cognitive abilities of the modeller in making inferences about the target system (Kuorikoski and Ylikoski 2015). Constructing a theoretical model is first and foremost a process that involves formalizing assumptions about the causal structure of the target in such a way that one can apply explicit inference rules to derive conclusions that would not be accessible without an external inferential aid of some kind. This is not a trivial point, but a crucial function of models in the study of complex systems. Given a target system composed of many components that simultaneously interact with each other, it is almost impossible to track the behaviour of the system very far just by 'running a simulation in one's head'. Constructing a model in which assumptions about the relations between the components are explicitly coded and the rules for deriving conclusions from those assumptions are tractable is immensely helpful and can reveal surprising results.

Consider again the Lotka-Volterra model. The model is built on an assumption that on the scale of an ecology, predation leads to an interdependency between the predator and prey population growth. The empirical support for this assumption comes from straightforward observations of predator and prey interactions; predators kill prey, prey abundance limits the number of predators. Constructing a mathematical expression of this assumption allows one to derive various conclusions about predator-prey systems, some of them surprising. A famous example is the Volterra principle, which states that if an external shock such as disease were to kill both predators and prey in equal proportions, one would see a subsequent reduction of the predator population in proportion to the prey population. This result is unexpected: it is not intuitively obvious that it follows from the simple qualitative fact that predator and prey populations mutually constrain each other, one needs to study the model to see that this is the case. But despite the phenomenological resemblance to empirical discovery—the experience of novelty or surprise upon learning the Volterra principle—this result is already implicit in the mathematical expression of the key assumption behind the model.

Instead of providing a new piece of evidence akin to an empirical

finding, the model here functions as a tool that enables inferences about the behaviour of the target based on limited assumptions about its causal structure. Studying the model thus provides potential insight directly into the behaviour of the target in virtue of showing what consequences are licensed by the assumptions encoded in the model. No intermediate step of comparing an abstract or fictional entity to a concrete target system is implicated in the process. What is learned through the derivation of the Volterra principle is a consequence of the modelling assumptions, and this result is as credible as tightly the assumptions are constrained by empirical evidence gleaned from the target itself. The inference is thus not ampliative in the way inferences based on evidence from animal models are.

These considerations hint at an alternative, inferential account of theoretical modelling. Viewed as a through-and-through inferential practice, theoretical modelling is a form of argumentation: a process of deriving conclusions from well specified premises through the application of explicit inference rules (Kuorikoski and Ylikoski 2015). In this view, models do not apply to their targets in virtue of a special type of mediated representation relation. Rather the model can be used to represent a target, in the sense of telling something truthful about it, insofar as the modelling assumptions are empirically adequate and the rules of deriving conclusions from the model are reliable.

When models are viewed as inferential aids, questions about the validity of a model become questions about the accuracy of the modelling assumptions and the reliability of the inferences drawn from those assumptions. Improving the analogy between the model and the target, such as adding terms for carrying capacity or intraspecific competition in the Lotka-Volterra model, does not involve comparing the target system to an abstract or fictional artefact specified by the model description. Instead, validity of the model depends on the degree to which the modelling assumptions, encoded in the formal expressions used in the model, are true of the target. More often than not, an advantage of constructing a formal model is that it forces the modeller to specify assumptions that were previously vague or ambiguous.

The role of robustness analysis appears in a different light as well. There is typically some slack between the evidence on which

modelling assumptions are based, and the formal expression of the assumptions as they are coded into the model. For example, knowing through observation that predator and prey species interact in mutually constraining ways does not itself determine a unique mathematical expression to describe the dependency between them. Also, deliberately unrealistic assumptions made for the sake of tractability are obviously not empirically supported. To evaluate the credibility of model-based inferences, one needs some method for identifying those conclusions that are stable through a range of alternative but equally credible formulations of the modelling assumptions. This is what robustness analysis of theoretical models does, by isolating conclusions that do not crucially depend on a specific set of modelling assumptions. Robustness analysis of theoretical models can thus be seen as a method for assessing the credibility of model-based inferences (Kuorikoski and Lehtinen 2009).

One can now distinguish two meanings for the notion of modelling. One refers to the use of a surrogate system in order to create evidence that is extrapolated to a different target system or class of systems. The other refers to the use of external representational media (graphs, equations etc.) to express assumptions about the causal structure of a system in a way that facilitates inferences about its behaviour or properties. Animal models fall squarely into the first mentioned role. A model organism is in a literal sense a surrogate for its intended target, studied to provide evidence of the target's features in an indirect way. Validating an animal model consists in providing support for the assumption that the model is similar to the target in the causal structure that generates the features/properties of interest, and of discharging possible errors due to known or unknown dissimilarities between the model and the target. The former task may involve direct comparisons to the target, circumstantial empirical evidence, deliberate manipulation of the model's features, or appeal to shared evolutionary origins. The latter task is often more difficult, but when several candidate model species are available, empirical robustness analysis can be applied to distinguish which behaviours are more broadly reproducible across heterogeneous models and which are dependent on idiosyncratic features of specific models.

The case of theoretical models is not as straightforward.

Theoretical models do provide the kind of inferential leverage described above, but might at times exhibit characteristics that are in practice akin to a genuine surrogate system; think of simulation models whose operation is in practice opaque to the modeller. However, even in such cases, the behaviour of the model is determined by the assumptions that the modeller has explicitly coded into the model. Thus even in the case of complex simulation models, the epistemic gain that the model provides is one of allowing the modeller to follow through the consequences of the modelling assumptions that the modeller is already aware of and can refine, should new evidence about the target require it. This is not the case with animal models, in which the dependencies that govern the behaviour of the model are not determined by the modeller, but are instead something that one wishes to learn by studying the model. Even when the strategies of validating models and the dynamics of studying them through subsequent rounds of (quasi-) experimental manipulation are similar, animal models and theoretical models serve different purposes.

7 Conclusions

I have argued that there is no clear cut difference between model organism research and theoretical modelling with respect to the strategies used to justify inferences from the model to the target. Although in many cases animal models in biology are validated through phylogenetic reasoning or circumstantial empirical evidence, researchers may just as well apply the strategies of controlling for disanalogy and robustness reasoning for validating animal models. However, the points raised by Levy and Currie are by no means of little importance. Levy and Currie describe the types of argument that biologists give for treating findings in one species as evidence of features of other species—the reasoning through which experimental organisms are awarded their status as surrogate systems from which evidence is extrapolated to other biological systems. I have merely pointed out that strategies that in Levy and Currie's taxonomy of models would be associated with theoretical modelling can be applied in animal model research as well. This does not falsify Levy and Currie's basic claim that theoretical models and model organisms serve different epistemic purposes. I have argued that this distinction

pointed out by Levy and Currie really has to do with the epistemic role or purpose of studying a model, rather than the strategies that are used to validate models.

A characteristic epistemic value gained by the use of theoretical models is the inferential leverage that formalizing one's assumptions about the target system gives for reaching conclusions that follow from those assumptions. To achieve this epistemic profit, no *further* assumption about analogy is needed to carry the modelling result over to the target system—the conclusions drawn with the help of a theoretical model are true of the target insofar as the modelling assumptions are true and the rules of inference applied are reliable. For this inferential purpose, adjusting the model specification in light of knowledge of the target's features is done to ensure that the modelling assumptions are empirically adequate, and robustness analysis is employed to check which modelling results are independent of any remaining unrealistic auxiliary assumptions of the model. In animal model research, similar justification strategies may be used for validating the model, but their purpose is to establish that the model is a credible surrogate source of evidence for hypotheses about the target system. The more similarities between the model and the target one can establish by whatever means, and the more secure one can be that one's results are not distorted by remaining dissimilarities, the better the model fits its role as a stand-in for the target. It is thus the role of a model as an inferential aid rather than a surrogate source of evidence that distinguishes theoretical models from model organisms.[4]

Veli-Pekka Parkkinen
University of Bergen
veli-pekka.parkkinen@uib.no

References

Anitschkow, Nikolai. 1913. Über die Veränderungen der Kanichnenaorta bei

experimenteller Cholesterinsteatose. *Beitrage zur pathologischen Anatomie und zur allgemeinen Pathologi* 56: 379—404.

Anitschkow, Nikolai. 1933. Experimental arteriosclerosis in animals. In *Arteriosclerosis*, ed. by Edmund. V. Cowdry. New York: The MacMillan Company, 271—322.

Ankeny, Rachel. 2009. Model organisms as fictions. In *Fictions in Science. Philosophical Essays on Modelling and Idealisation*, ed. by Mauricio Suárez. London: Routledge, 194—204.

Ankeny, Rachel; and Leonelli, Sabina. 2011. What's so special about model organisms? *Studies in the History and Philosophy of Science* 41: 313—23.

Ankeny, Rachel; Leonelli, Sabina; Nelson, Nicol C.; and Ramsden, Edmund. 2014. Making organisms model human behavior: situated models in North-American alcohol research, since 1950. *Science in Context* 27(3): 485—509.

Aschoff, Ludwig. 1933. Introduction. In *Arteriosclerosis,* ed. by Edmun.V. Cowdry. New York: The MacMillan Company, 1—19.

Bolker, Jessica A. 2009. Exemplary and surrogate models: two modes of representation in biology. *Perspectives in Biology and Medicine* 52(4): 485—99.

Bruger, Maurice; and Oppenheim, Elliot. 1951. Experimental and human atherosclerosis: possible relationship and present status. *Bulletin of the New York Academy of Medicine* 27(9): 539.

Duff, G. Lyman;**d** McMillan, Gardner C. 1951. Pathology of atherosclerosis. *The American Journal of Medicine* 11(1): 92—108.

Frigg, Roman. 2006. Scientific representation and the semantic view of theories. *Theoria* 55: 37—53.

Giere, Ronald. 2004. How models are used to represent reality. *Philosophy of Science* 71: S742—52.

Godfrey-Smith, Peter. 2009. Models and fictions in science. *Philosophical Studies* 143: 101—16.

Hirsch, Edwin F.; and Weinhouse, Sidney. 1943. The role of the lipids in atherosclerosis. *Physiological Reviews* 23(3): 185—202.

Katz, Louis N. 1952. Experimental atherosclerosis. *Circulation* 5(1): 101—14.

Kuorikoski, Jaakko; and Ylikoski, Petri. 2015. External representations and scientific understanding. *Synthese* 192(12): 3817—37.

Kuorikoski, Jaakko; and Lehtinen, Aki. 2009. Incredible worlds, credible results. *Erkenntnis* 70(1): 119—31.

Leary, Timothy. 1944. The role of lipids in atherosclerosis. *The Journal of American Medical Association* 124(6): 358—86.

Levy, Arnon; and Currie, Adrian. 2015. Model organisms are not (theoretical) models. *The British Journal for the Philosophy of Science* 66(2): 327—48.

Maugeri, Paol B.; and Blasimme, Alessandro. 2011. Humanised models of cancer in molecular medicine: the experimental control of disanalogy. *History and Philosophy of the Life Sciences* 33: 603—22.

Morrison, Margaret. 2009. Fictions, representations and reality. In *Fictions in Science. Philosophical Essays on Modelling and Idealisation*, ed. by Mauricio Suárez. London: Routledge, 110—35.

Morgan, Mary S.; and Morrison, Margaret. 1999. *Models as Mediators: Perspectives on Natural and Social Science*. Cambridge: Cambridge University Press.

Mäki, Uskali. 2005. Model are experiments, experiments are models. *Journal of Economic Methodology* 12(2): 303–15.

Nelson, Nicole C. 2013. Modeling mouse, human, and discipline: epistemic scaffolds in animal behavior genetics. *Social Studies of Science* 43(1): 3–29.

Peters, John P.; and VanSlyke, Donald D. 1946. *Quantitative Clinical Chemistry*. Baltimore: Williams & Wilkins.

Weber, Marcel. 2005. *Philosophy of Experimental Biology.* Cambridge: Cambridge University Press.

Weisberg, Michael. 2004. Robustness analysis. *Philosophy of Science* 73(5): S730–42.

Weisberg, Michael. 2013. *Simulation and Similarity: Using Models to Understand the World*. New York: Oxford University Press.