

1 Introduction

Our dependence on other species in biomedical research and medicine calls for philosophical scrutiny. Several hundred million animals are used worldwide each year in preclinical research and for drug testing, with mice and rats alone counting about 120 million animals (Taylor and Alvarez 2019; Carbone 2021; Cait et al. 2022).¹ Animal models serve as proxies for human diseases in a basic research and drug development, but the benefits and limitations of animal models are contested topics in science, bioethics, and public debates. How – and to what extent – can we learn about humans by studying and experimenting on animal models? What functions do animal models play in biomedical research, and why are specific animals chosen for specific purposes? How are boundaries between humans and animals constructed and negotiated in this process, socially and experimentally, and what are the main translational challenges and ethical concerns? And, given persistent problems of translational failures, what are the prospects of replacing animal models with animal-free methods in the future? This Element delves into these questions and highlights the intertwinement of epistemic, practical, and ethical issues faced in translational research. I hope this Cambridge Element will raise questions of interest to philosophers, social scientists, and scientists alike.

Another Cambridge Element on *Model Organisms* already covers many of the philosophical implications of the use of animal models in the life sciences (Ankeny and Leonelli 2020). Why this additional Element on *Animal Models of Human Disease*? This Element is focused on animal models used in the context of translational models, that is, on animal models that are *human-directed*. While many animal models in translational research are *model organisms* in the sense defined by Ankeny and Leonelli, the two terms should not be conflated. Model organisms are non-human organisms that are standardized to display general genetic or physiological features, and they thus have a broad representational scope and institutional support structure that allow for cross-species knowledge integration (Ankeny and Leonelli 2011; Leonelli and Ankeny 2012). Because model organisms are used to study general physiological features across a variety of species, their role as models cannot be reduced to the epistemic interest in improving human health. For example, the reliance on the thale cress, *Arabidopsis thaliana*, as a model organism in plant biology does not hinge on the relevance for human health (Leonelli 2007).²

¹ The numbers are estimated from registers of ethics approvals, including only animals that are *legally* considered to have moral status, i.e., it does not include most invertebrate species. See Sections 5 and 6 for further discussion.

² This, however, does not leave out the possibility that studies on *Arabidopsis thaliana* can inform human genomics and medicine (e.g., Jones et al. 2008).

Thus, the focus of this Element is in this sense narrower in zooming in on how and why animals, or parts of animals, are used for the specific purpose of learning about human diseases or improving human health (Huber and Keuck 2013).³ But the focus is also broader in the sense that not all human-directed models are subsets of model organisms. As illustrated in this Element, translational animal models take on a variety of epistemic roles that are worth exploring as separate topics.

Animal models of human disease do not necessarily represent a large class of organisms, or even a large group of humans. They may, for example, serve as *surrogate models* that “stand in” for specific patient groups or even specific patients in drug development or drug testing (Bolker 2009; Green et al. 2021). Some animal models are also investigated as what some scientists call *negative models* because their resistance to human diseases or pathological conditions is medically relevant (Green et al. 2018). Moreover, animals can play more *instrumental roles* in medicine as diagnostic tools or detection devices that are not easily accounted for through a standard account of scientific models as representations of targets (Germain 2014; Knuuttila 2021). Scrutinizing the various functions of animals in biomedical research can therefore extend and deepen philosophical discussions on modeling in general. Translational research can also offer insights into the epistemic challenges of balancing model virtues that *represent* and *reduce* the complexity of the target systems. In this context, standardization of models to improve the reproducibility of results in the laboratory can be counterproductive if the aim is to represent the complexity and variation encountered in the clinic. Practices of animal modeling thus raise fundamental questions about what constitutes good evidence in science and medicine.

Zooming in on specific uses of animals as means for improving human health also exposes important questions about how we relate – physiologically and emotionally – to other species (Sharp 2019; Kiani et al. 2022). The physiological and behavioral similarities between humans and non-human animals simultaneously facilitate translational inferences and produce ethical conflicts. In this sense, epistemic questions about model validity are intertwined with ethical considerations on the weighting of animal welfare and human interests (Singer 1975; Regan 1983). Are such considerations inescapable tensions in animal research, or can caring for experimental animals be reconciled with objectivist norms for good science? Does animal experimentation stabilize the distinctiveness of the human or remind us of the relatedness to other animals?

³ I do not have space to discuss the relationship between human and veterinary medicine. But transfer of knowledge and drug development across these contexts exemplify how some animals can benefit from human medicine – and vice versa (Alder and Easton 2005).

This Element suggests that the methodologies employed in human-directed animal modeling can be a prism through which our understanding of the human is refracted, raising fundamental questions about what defines human nature in comparison to non-human species (Efstathiou 2019; Svendsen 2022, see also Ramsey 2013; 2023).

In zooming in on these issues, the future of animal models must also be critically scrutinized. The perceived necessity and adequacy of animal models in biomedical research and medicine are increasingly contested issues. Critics have for decades stressed how differences between species can lead to misleading inferences, especially when animal models are used to predict the efficacy and adverse effects of drugs (LaFollette and Shanks 1993; 1995). The concerns have been growing with recent studies documenting highly varied translational success of animal modeling (Mullard 2016; Striedter 2022; Swaters et al. 2022). As Leenaars et al. (2019) observe, discussions in the scientific field currently revolve around two main perspectives with different implications for the future of animal models: one explaining the translational failures by suboptimal experimental design and calling for improvements in animal research, and another calling for a radical shift to non-animal methods.

Since the Federal Drug and Cosmetics Act of 1938, animal testing has been a requirement for the protocols of drug development to ensure the safety and efficacy of drugs before they can enter first-in-human trials. But we may currently be witnessing significant changes. In September 2021, the European Parliament almost unanimously voted for an action plan to phase out animal experimentation for research and drug testing (Marshall et al. 2022).⁴ The same year, the American Congress passed a bill called the FDA Modernization Act 2.0, which was signed by President Biden in December 2022. The Modernization Act 2.0 removes the strict requirement of animal testing and allows drug developers to use alternative nonclinical tests in drug development.⁵ Although it is beyond the scope and purpose of this Element to cover the political and public debates on animal research, the intensified focus on reducing animal experimentation calls for a better understanding of how animal models are used in translational research – and what alternatives there may be for replacing or reducing these. My aim is not to defend a specific view on the future of animal models or the philosophical interpretation of animal models in general, but rather to unpack core questions

⁴ The aim to phase out animal models in biomedical research was already part of the EU Directive 2010/63, stating that “wherever possible, a scientifically satisfactory method or testing strategy not entailing the use of live animals shall be used” (see also Smith et al. 2013 and Section 6). For more information on the recent action plan, see: www.europarl.europa.eu/news/en/press-room/20210910IPR11926/meps-demand-eu-action-plan-to-end-the-use-of-animals-in-research-and-testing.

⁵ [www.congress.gov/bills/117th-congress/senate-bill/5002](https://www.congress.gov/bills/117/congress/senate/bills/5002), accessed December 20, 2023.

and considerations that I find important for a nuanced debate. I hope that the Element can provide a conceptual framework to articulate the diversity of epistemic and practical functions of translational models, as well as some of the challenges and proposed solutions in translational research. My aim is also point to questions that I find intriguing but that have not yet received much philosophical attention.

The Element is structured as follows. Section 2 provides introductory reflections on how and whether animals can be said to serve as models in translational research, considering also how animals are constructed or engineered for specific translational purposes. Section 3 explores different aspects of the persistent tension between standardization and variation of animal models. This includes an introduction to the important role of standardized model organisms in translational research but also to the philosophically intriguing roles of non-canonical organisms, including so-called “negative models.” With this background, Section 4 revisits the virtues of animal models when these act as “patient substitutes” and stand in for human patients in ways that sometimes blur boundaries between animal model and human patient. Section 5 discusses the epistemic roles of animals in biomedical research that may go beyond the traditional focus on representation in philosophical discussions on models. This involves the temporality of model development and the use of animals as diagnostic tools or as material and collaborative resources. Section 6 discusses the future of animal models, including the potentials and challenges of replacing animal models with non-animal methods, such as in vitro models based on human cells. Finally, Section 7 summarizes the key points and ends with concluding reflections on the need for further philosophical work on the topic of animal models of human disease.

2 Animals as Models of Human Disease

The best material model for a cat is another, or preferably the same cat.

- Rosenblueth and Wiener (1945)

Paraphrasing Rosenblueth and Wiener’s famous quote, one might say that in translational research “the best material model of a human is another human, or preferably the same human.” In both cases, however, one would misunderstand what a model is. In the broadest sense, a model is a simplified representation of a system or phenomenon that is used to understand, predict, or simulate a real-world behavior or relationship. A more detailed or representationally realistic model is not always better. Rosenblueth and Wiener illustrate this point by referencing Jorge Luis Borges’ (1954/1972) fictive story on “exactitude in science,” where the science of cartography reaches the highest level of

perfection and maps become as big and complex as the landscapes they represent. Such maps are useless because their exact accuracy prevents them from performing their epistemic function as a map, that is, as a simpler overview that helps us navigate in complex spaces. Rosenblueth and Wiener's paper is about the role of theoretical models in science. But animal models can similarly "mediate" between our theoretical understandings of disease mechanisms and a real-world target (Morrison and Morgan 1999) by allowing for more practically accessible or ethically permissible experimental interventions on causal mechanisms. How should the role of animal models be understood in comparison to theoretical models? And are researchers confronted with a similar tension between representing and reducing the complexity of the target? Let us take a closer look at the characteristics of animal models in translational research.

2.1 Modeling Human Disease by Intervening on Animals

The comparison between theoretical models and animal models can be misleading in the sense that we may overlook what makes experimenting on animals special. Levy and Currie (2015) argue that model organisms are not (theoretical) models because they are "samples from, or specimens of, a wider class" (p. 328). Inference from model organisms, they argue, are not made merely through artificially constructed and abstract analogies between model and target. Rather, model organisms are special in providing circumstantial or *phylogenetic* evidence, as members of the same phylogenetic class under investigation (see also Love 2007; Steel 2008; Weber 2005). From this perspective, learning about human diseases by intervening on animals is grounded in the evolutionary conservation of phenotypic traits based on homologous genes and "elementary building blocks" that are universally shared among many organisms (Changeux 2006). Inferences are justified not through idealized approximations but through insights into basic causal features that many organisms have in common. For example, many mechanisms regulating embryonic development appear to be evolutionarily conserved across many species, thus justifying why the neural circuits of an invertebrate such as *Caenorhabditis elegans* can serve as a simple model for neurological disorders in humans (Schaffner 2001). According to Weber (2001; 2005), inferences from such reduced models are justifiable, because biological mechanisms are *hierarchically structured* such that lower-level mechanisms are typically similar across species, even if higher-level capacities differ. From this perspective, extrapolation from interventions on lower-level mechanisms in a different species can be justified if relevant difference-makers can be documented in both contexts.

Undoubtedly, phylogenetic relatedness is important for understanding how interventions on animals can be informative for medicine. Nevertheless, one should be careful not to commit what LaFollette and Shanks (1995) call the modeler's *phylogenetic fallacy*, referring to the uncritical assumption that phylogenetic continuity implies underlying causal similarity. LaFollette and Shanks argue that phylogenetic relatedness cannot justify the use of animal models as *causal analog models*, as phylogenetic relatedness does not justify direct causal inferences. In their view, evolutionary conservation of physiological traits can only support the use of animal models as *hypothetical analog models* to suggest possible mechanisms for further investigation. Indeed, evolutionary conservation of many basic mechanisms does not always extend to homologous links between genes and disease mechanisms or drug metabolism, for example, when comparing the evolution of gene networks and molecular mechanisms in humans and mice (Perlman 2016). Moreover, prior knowledge of disease-relevant causal mechanisms is often not available to guide this exploration of lower-level mechanisms. As Baetu (2016) highlights: "in the initial stages of [translational] research, relying on similarities at the level of the causal structures is of little use, since it is precisely these structures that researchers aim to elucidate" (p. 10). This challenge is sometimes called "the extrapolator's circle" (Steel 2008). In preclinical modeling, the epistemic uncertainty of the translational models is often intertwined with ontological uncertainties about what features of the human disease are most relevant to recapitulate in the model (Green et al. 2022).

The most suitable model must be evaluated through iterative steps, involving not only structural and functional similarities of shared phylogenetic factors or molecular mechanisms but also investigations of what Baetu (2016) calls "symptom similarity," understood as phenotypic features linking animal models to translational targets in experimental interventions. Focusing on symptom similarity can also reveal how the best translational model is not always the phylogenetically closest relative. Chimpanzees are the closest living relatives to humans, with an astonishing 99 percent overlap in protein-coding genetic sequences (Suntsova and Buzdin 2020). Yet, despite the high degree of genetic and physiological similarity, AIDS research in the 1980s and 1990s was confronted with the difficult challenge that HIV-infected chimpanzees did not develop the AIDS-related symptoms seen in humans (van Akker et al. 1994).⁶

⁶ AIDS-like symptoms and increased mortality have later been observed in wild chimpanzees infected with versions of simian immunodeficiency viruses (Keele et al. 2009). While this finding challenges previous conclusions on species-specific immune adaptations, the example still calls for caution concerning cross-species extrapolation, even when the animal is a "close relative." Chimpanzees have been important animal models in vaccine development (e.g., hepatitis), but invasive research on chimpanzees is now largely prohibited due to ethical concerns (Harding 2017).

These challenges, alongside ethical concerns of using chimpanzees for research, have made researchers explore other animal models, including macaque monkeys infected with simian immunodeficiency viruses, cats infected with feline immunodeficiency virus, and rodent models “humanized” through transgenic techniques to resemble human immune responses. The example of AIDS research thus illustrates how several animal models are often needed, each contributing with some pieces of information to a “mosaic description” of disease mechanisms (Baetu 2016; see also Green 2013; Baetu 2014). Both theoretical and animal modeling therefore involves what Rheinberger nicely formulates as the process of “shuttling back and forth between different spaces of representation” (Rheinberger 1997, pp. 108–109).

Another observation that challenges the strong reliance on justification of model choice via phylogeny is that translational models are no longer limited to the organism’s evolutionary features but are often genetically modified to minimize disanalogies to human targets (Maugeri and Blasimme 2011). For example, genetically engineered mouse models (GEMMs) are used to study a variety of diseases including Parkinson’s and Alzheimer’s disease, Down’s syndrome, rheumatoid arthritis, obesity, and diabetes, just like genetically modified porcine models are important animal models in organ and tissue transplantation research (Huber and Keuck 2013; Hardesty 2018; Lowe 2022). According to Parkkinen (2017), it is therefore not possible to distinguish the epistemic strategies of animal and theoretical models with reference to the role of phylogeny alone. Yet, he stresses that this does not challenge the basic claim of Levy and Currie (2015) that theoretical models and animal models serve different epistemic purposes. Parkkinen suggests that a distinction should instead be drawn between theoretical models as *inferential aids* and animal models as *surrogate sources of evidence* (Parkkinen 2017). Drawing on Bolker’s (2009) notion of surrogate models (discussed further in Section 4), Parkkinen argues that animal models serve as material surrogates for human patients, making the degree of (material) similarity between model and target more pressing in this context. He contends that: “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 first its role as a stand-in for the target” (p. 496). Indeed, animals are used as models in biomedical research because they are considered sufficiently *biologically like* human counterparts to warrant causal inferences (Lewis et al. 2013), and yet sufficiently *morally different* from humans (Svendsen and Koch 2013). But the notion of “similarity” can be defined in different ways, and what constitutes relevant or sufficient biological similarity (and moral worth) depend also on the historical context and the epistemic purpose of specific studies.

The latter point can be illustrated through a scientific discussion in epilepsy research. In a comment on a study of spontaneous seizures in “epileptic rats” (Nissinen and Pitkänen 2007), Mazarati (2007) distinguishes between what he calls the “analogical modeling approach” and the “conceptual modeling approach.” Analogical modeling stresses the representational matching of model and target, akin to what Parkkinen (2017) hints at. From this perspective, the best model to study human epilepsy would be a rodent model *of* epilepsy, that is, a rodent that maximally represents the symptoms and symptom development in the human counterpart (e.g., spontaneous seizures). Conceptual modeling, in contrast, emphasizes that models should not merely resemble targets but should provide easier experimental access to causal factors that cannot be studied without distorting and simplifying the phenomenon of interest (e.g., experimentally induced seizures). Ratti (2020) similarly distinguishes between the notions of “models of” and “models for,” where the latter denotes how some models are chosen not because of their direct representational or explanatory force, but because of the interventionist strategies they allow for. In the case of epilepsy research, Mazarati (2007) stresses that “a key rationale underlying the conceptual model is to establish logical relationships among variables rather than simply to account for as many variables as possible. Idealization is a key feature of the conceptual model, allowing for simplification of the phenomenon to such an extent that it can be studied effectively” (Mazarati 2007, p. 112). Mazarati thus points to a relationship between model idealization and practical efficiency (or interventional relevance), not unlike what has also been discussed for (some) theoretical models as “minimal models” (Batterman and Rice 2014).

Mazarati views the conceptual approach to models as superior, but there may be benefits to using both types of models and avoiding generalizations about what constitutes a good translational model, at least if the question is addressed in isolation from specific research questions. A focus on the *validity of the inference* from animal models, given specific aims, may be more fruitful than focusing on the model’s similarity to the target. It is common in translational research to distinguish between a model’s (i) face validity, (ii) construct (or target) validity, and (iii) predictive validity (Denayer et al. 2014; Lemoine 2015; see also Striedter 2022, p. 21). *Face validity* is emphasized in what Mazarati (2007) calls the analogical modeling approach which emphasizes the similarity of phenotypic traits or symptoms in the model and target “on the face of it.” *Construct (or target) validity* refers to similarity relations in the underlying causal mechanisms of a disease-relevant process in a model and a target, which can help explain why a disease occurs or a treatment works