2 PHILOSOPHY OF MODELS IN BIOLOGY

> All models are wrong, but some models are useful. So the question you need to ask is not "Is the model true?" (it never is) but "Is the model good enough for this particular application?"
>
> —BOX ET AL. (2009), P. 61

When scientists say they are using a model, they mean that they are examining one thing, the model, in hopes of better understanding not just the model itself but the modeled system—the target. As the prominent statistician George Box articulated in the chapter's opening quotation, models of this kind are never identical to the target. However, they should be analogous to the target in some respects and more accessible to examination (Oreskes, 2007). Moreover, the analogies should be more than superficial. In particular, good models should resemble the target at a causal level so that the responses of the model to experimental manipulations can generate hypotheses about how the target should respond to similar interventions (Hesse, 1963; LaFollette & Shanks, 1995; Shapiro, 2004). If these predictions are confirmed, scientists will have learned something about the target; if they fail, then modifications of the model are called for.

Most philosophers and scientists probably agree on the preceding statements (e.g., see Creager et al., 2007), but scientists use many different kinds of models, and philosophers have offered a variety of classification schemes for them. For example, they have distinguished "theoretical models" from "model organisms" (Levy & Currie, 2015; Parkkinen, 2017), "exemplary models" from "surrogate models" (Bolker, 2009), "models of" from "models for" (Keller, 2000), and "model organisms" from other organisms that experimental biologists may study (Ankeny & Leonelli, 2011). The lack of perfect correspondence between these distinctions likely stems from the fact that different philosophers have emphasized different kinds of biological research, which differ in their aims. Particularly important is that some biologists try to extrapolate from their experimental

model to a very limited set of targets, while others seek broad generalizations or even universals (Ankeny & Leonelli, 2011; Parkkinen, 2017). This variation in model usage and philosophical distinctions is important to recognize, but difficult to review without going into more details than this book can accommodate (or its intended audience would likely tolerate). Interested readers may consult the cited references to gain entry into this stimulating literature.

For present purposes, it is sufficient to distinguish between *abstract models*, which are purely conceptual, and *material models*, which in biology comprise mainly cultured cells, plants, and animals (Rosenblueth & Wiener, 1945; Pease & Bull, 1992). Although biologists increasingly use both types of models, the bulk of their research involves material models that are studied as proxies for one or more target systems (Bolker, 2009). Some biologists use such material models to improve livestock or agricultural products, but this book focuses mainly on the application of material models to human health and disease, which is why model plants (e.g., maize and *Arabidopsis thaliana*) are given short shrift. The book also neglects the use of material models to better understand nonhumans, which is central to veterinary medicine (Cunningham et al., 2010).

2.1 ABSTRACT MODELS

Abstract models mimic key elements and relationships of the target system and, importantly, do not include extraneous detail. Paradigm examples of abstract models are mathematical models, which typically contain multiple differential equations. Such models are not meant to be complete representations of the target system; rather, they include only those factors and relationships that are thought to be necessary for explaining the target system's behavior. If subsequent testing reveals that the model's behavior fails to mimic that of the target in some important respects, then the model should be modified.

This is, in fact, the principal purpose of abstract models: they reveal deficiencies in our thinking about the target and, therefore, require us to modify our assumptions (Gunawardena, 2014). In addition, abstract models can drive discovery by generating novel predictions that are then confirmed in the target system, but were previously unknown (Hesse, 1963). Although abstract models are ideally framed in the language of mathematics, they may also take the form of less precise conceptual models (Thompson, 1917). However, all abstract models are constructed to be as simple and elegant as possible, but no simpler (as Einstein supposedly once said).

Mathematical models have long been central to physics, but they were slow to take hold in biology. For example, the mathematician and early computer scientist Alan

Turing in 1952 published a mathematical model to explain the formation of structural patterns in organismal development, but his model was ignored by biologists for many years, partly because of questionable assumptions (Keller, 2002). More successful was Hodgkin and Huxley's (1952) mathematical model of the neuronal action potential using experimental data obtained from squid giant axons.

Mathematical models are still relatively rare in biomedicine, but pharmacology has long emphasized mathematical models (van der Graaf et al., 2016), and quantitative models have recently been developed for several aspects of cancer biology (Altrock et al., 2015). Abstract models that do not employ mathematics have, of course, been common in all realms of biology for many years, even if those conceptual models were sometimes rather fanciful (Arikha, 2007).

2.2 MATERIAL MODELS

As defined here, material models are concrete, physical entities that are analogous to the target system in important respects but have numerous properties that are not shared with the target and cannot easily be stripped away. The history of geology, for example, includes various material models of the earth's crust that were built from layered materials that could be compressed laterally to see if they would produce folds analogous to mountain ranges. These models "made the inaccessible accessible" (Oreskes, 2007).

Staying within geology, mathematical models of how the earth's magnetic field is generated from interactions between the earth's rotation and subterranean currents of molten metal are too complex to solve, and the entire earth is obviously too large for controlled experiments. Therefore, geophysicists have created a 1-meter diameter, rapidly spinning metallic sphere filled with liquid sodium to study its induced magnetic field (Lathrop & Forest, 2011). Among other things, they hope to discover whether the magnetic field of this material model will occasionally flip polarity, as the earth's field regularly does. In this example, the physical properties and dynamics of the material model were selected to match those of the earth closely, but differences clearly persist; the extent to which they are negligible remains to be seen.

Biologists likewise use material models to make the relatively inaccessible accessible. Most obviously, they perform experiments on animals that would be unethical to carry out on humans. In addition, biologists seek principles that are conserved across a large number of species. Of course, they cannot study all the earth's species and must, therefore, focus their studies on a select few, easily accessible species, which then function as representatives for a larger taxonomic group (Ankeny & Leonelli, 2011). The selected species are generally easy to breed and maintain in laboratory environments, and experimentally tractable. As the Nobel Prize–winning physiologist August Krogh

put it in 1929, some species are more "convenient" than others for experimental investigation. In fact, Krogh exhorted experimentalists to consult zoologists to help them find the most convenient species for whatever problem they are studying. Hans Krebs (1975) later referred to this observation and advice as "the August Krogh principle" (for more on this, see section 2.6 and chapter 3; for a list of all winners of the Nobel Prize in Physiology or Medicine, as well as their principal model systems, please consult the appendix).

A very general problem with the use of model species in biology is that evolution can modify any aspect of a species, which means that no one species in any particular study is a priori guaranteed to be representative of any (or all) other species, even at the molecular and cellular levels. Biologists may work with "simple animals" in the hope that those models exhibit the target's key elements and interactions without additional complexities, but evolution is not always conservative. Indeed, supposedly simple animals are often more complex than people imagine, and their complexity may be quite different from our own (Sterling & Laughlin, 2015). An analogous problem complicates in vitro research: cell and tissue culture systems inevitably differ from their in vivo counterparts in multiple respects. As critics of cell culture work like to proclaim, *in vivo veritas!* (Matarese et al., 2012).

Scientists can modify their animal and cellular models to make them more similar to the target, but it remains essentially impossible to strip away from such models all their extraneous, idiosyncratic features. The closest approximation to a stripped-down material model in biology is the recently created "synthetic bacterium" whose genome contains only essential genes that are conserved across bacteria (Hutchison et al., 2016). However, such a minimal organism is not a good model for research on complex human diseases.

2.2.1 Model Scope and Validation

An important aspect of material models in biology is that they vary considerably in scope: some serve as a model for a very small subset of organisms (e.g., only humans), whereas others are meant to represent taxonomically broad swaths of species, including potentially all life. Unfortunately, as noted in the previous section, it is impossible to know a model's scope ahead of time. Only after examining the target do we know whether our model selection served our purpose. This dilemma, which has been called the "extrapolator's circle" (Steel, 2007), is sometimes cited as an argument against all animal experiments, but this conclusion would be warranted only if findings obtained in animals predict human results at levels near or below chance.

As we discussed in chapter 1, the high failure rate of clinical trials is certainly concerning, but animal research has, over the long course of history, made an undeniably

positive contribution to medical progress (see chapters 5 and 6). Therefore, biological models have clearly sometimes been useful (or, as George Box would say, good enough for their particular applications). The real problem is that a model's utility is an empirical question and cannot, therefore, be assumed a priori (Levy & Currie, 2015). Moreover, even models that have proven to be useful for some research questions may fail when they are used to address other, different questions. In short, model systems can never be validated once and for all.

Within the field of biological psychiatry, researchers have proposed a multidimensional set of model validation criteria (McKinney & Bunney, 1969; Willner, 1984; Belzung & Lemoine, 2011). *Face validity* specifies that the model and its target should be similar "on the face of it," which in practice means that they ought to exhibit similar symptoms. By contrast, *construct validity* refers to a similarity of underlying causal mechanisms. The distinction between symptoms and mechanisms can sometimes be difficult to draw, especially for so-called biomarkers, but the distinction is valuable because models with strong construct validity may sometimes be useful even if they exhibit low face validity (van der Staay, 2006). The third main criterion is *predictive validity*, which refers to how well a model predicts the target's response to established or, ideally, new therapies (e.g., Fossat et al., 2014). Some models with high predictive validity may be little more than assays to find new drugs that are similar to those already known to be effective in the target system—a phenomenon dubbed "receptor tautology" (Geyer et al., 2012)—but the generation of successful predictions is clearly desirable for any model.

Although these three validity criteria have thus far failed to resonate beyond their field of origin, analogous criteria are likely used implicitly by most biologists as they evaluate material models in their research area. In general, construct and predictive validity seem more useful than face validity, mainly because assessments of the latter are often superficial or downright fanciful (Garner, 2014).

As mentioned previously, a major constraint on all material models is that they may harbor unsuspected, uncontrolled features that break their analogy to the target and thus limit their utility. This is true even for models that satisfy multiple validity criteria. Moreover, model selection is heavily influenced by factors that are rarely considered publicly, such as ethical concerns, costs, regulations, and societal pressures. Indeed, Dietrich et al. (2020) have identified a set of 20 interacting criteria that biologists typically consider, to varying extents, in selecting their models. Therefore, determining the "best" model for any particular question is usually a highly personal and subjective affair, heavily influenced by scientific tradition. This may be unavoidable, but the process would likely benefit from more extensive discussions (see chapter 7).

2.2.2 Model Modification

Biologists often try to mitigate the limitations of their material models by modifying their research organisms. For example, they frequently "standardize" their animals by creating highly inbred lines (see chapter 3). Although inbreeding tends to yield reduced fertility and other abnormalities in the short term, strong artificial selection over many generations of inbreeding can create viable lines that are genetically uniform. The reduced variability of these inbred strains makes it much easier to obtain statistically significant effects when the organisms are manipulated experimentally and compared with controls.

In addition, biologists often create modified strains in which one or more genes have been mutated, knocked out, or inserted into the genome. Once these genetically modified strains have been created, they can be maintained in central repositories and shipped to individual investigators as needed. The creation of such mutant libraries is a major reason why some species—notably mice, rats, yeast, and zebrafish—have become so widely used that they are sometimes called "model organisms" (Ankeny & Leonelli, 2011, 2020; Preuss & Robert, 2014; Katz, 2016) (figure 2.1). Researchers also tend to modify the genes of cultured cells, which is why cell lines in which such modifications are relatively easy to perform have become widely used (see chapter 4).

Aside from manipulating the genomes of their models, researchers often manipulate the environment in which those models live. For instance, the solutions in which cultured cells are grown (i.e., the media) can have profound effects on how the cells proliferate and develop. Similarly, raising animals in small cages with monotonous food is quite different from raising them in physically, socially, and nutritionally enriched environments, which in most cases more closely resemble the animals' natural state. For many studies it also matters whether the animals are raised in an artificially sterile environment or exposed to immune system challenges and allowed to grow their own intestinal microbiome. Even the temperature at which animals (or cells) are kept can have wide-ranging effects. For example, mice kept at the standard housing temperature of ~22°C, which is human room temperature, are chronically cold-stressed (Kokolus et al., 2013; Fischer et al., 2019). Although such environmental factors often receive less attention than genetic manipulations, they can have profound effects (figure 2.2) and should be considered integral to any animal or in vitro model.

Biologists sometimes study diseases that occur naturally in their research animals, but more frequently they modify the organisms to induce the disease. They may do so via genome manipulation or by means of other modifications that trigger processes thought to be involved in generating the disease (see chapter 5). The advantage of these artificial disease inductions is that the manipulated animals tend to develop the disease reliably and with relatively uniform symptoms. As in the case of inbreeding,

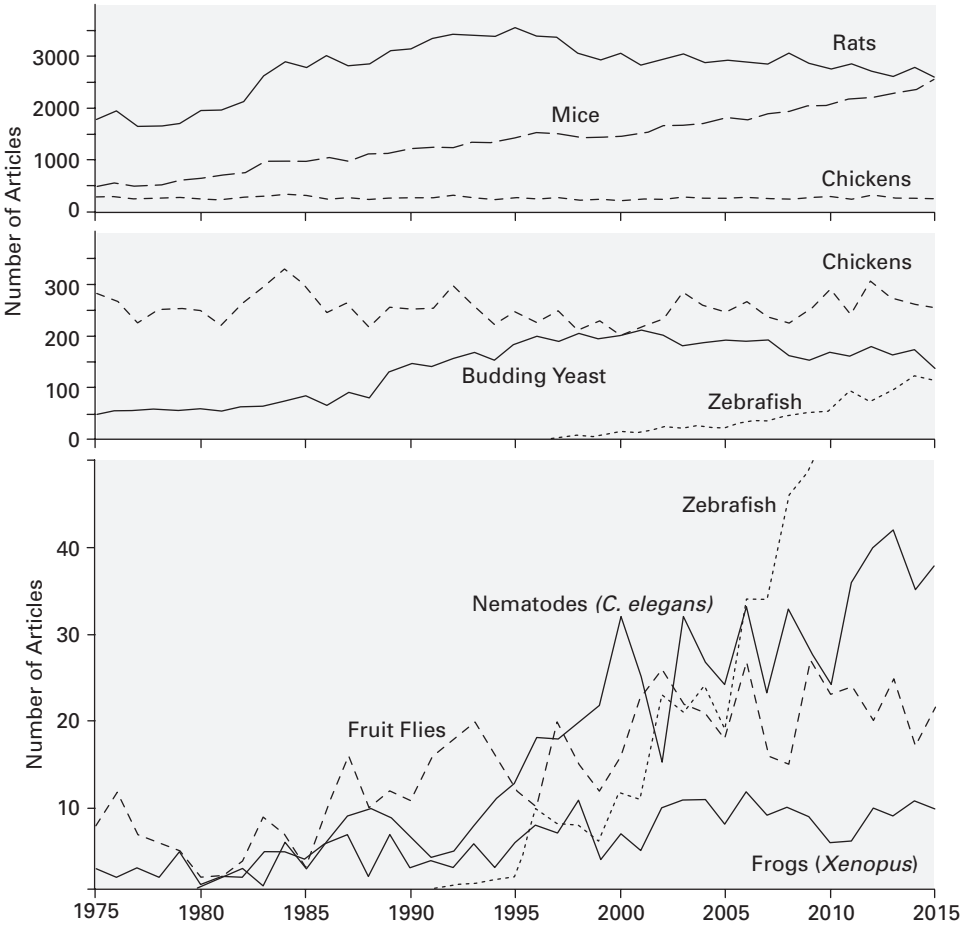


Figure 2.1

Variation in the popularity of select model organisms. Peirson et al. (2017) analyzed more than 3 million abstracts in the PubMed database for mentions of the principal National Institutes of Health–designated model organisms (spread evenly across 26 disease categories). The three graphs are from the same analysis, but show the data at different scales. Shown here are only data for the top eight of the 13 models in the analysis. Adapted from Peirson et al. (2017).

this reduction in interanimal variability makes it much easier to obtain statistically significant results in subsequent experiments, including tests of potential therapies. However, this research strategy assumes that experimenters have correctly identified what causes the human disease and that their manipulation triggers a cascade of disease processes that mimics human disease progression. If this assumption is false, then there is no guarantee that reversing the effects of an experimenter's manipulation will lead to therapies that are effective in humans. As we shall see in later chapters, this possibility should not be ignored.

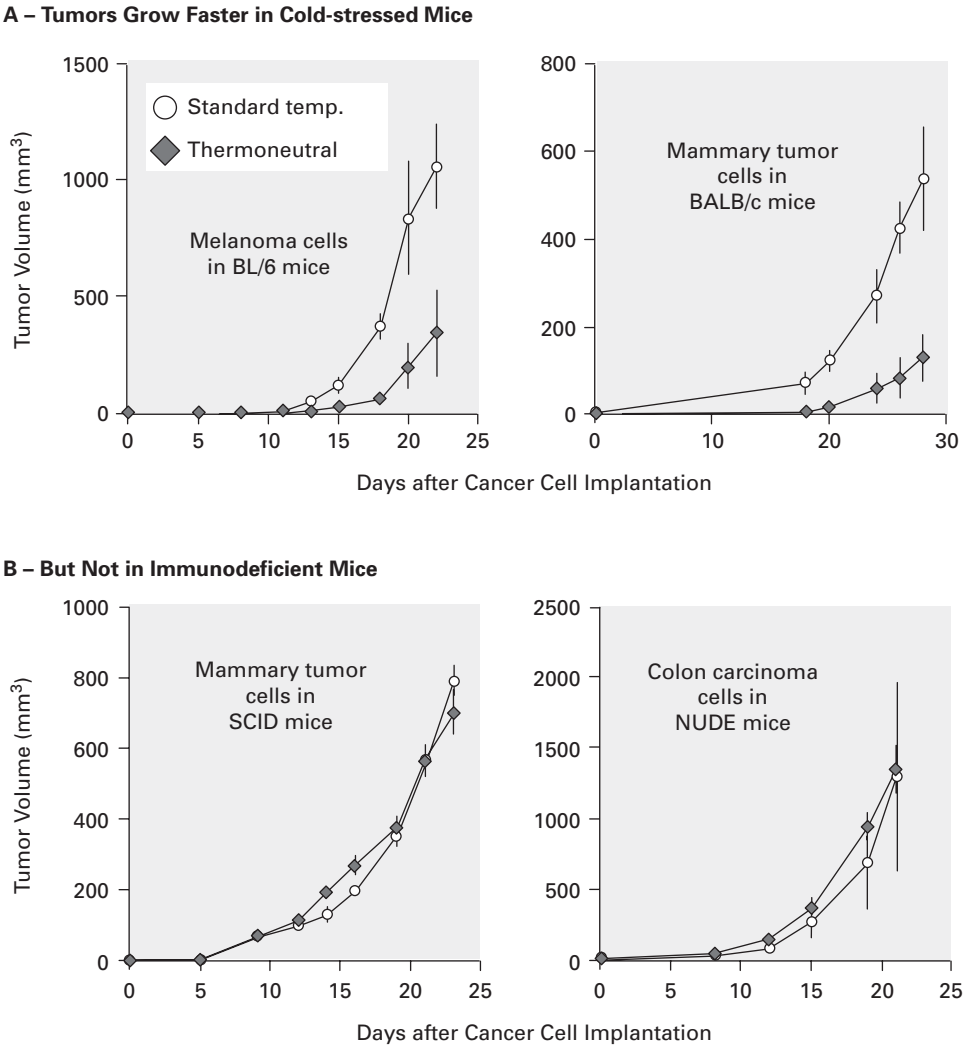


Figure 2.2

Mice are cold-stressed at temperatures that humans find comfortable. (*A*) Tumors grow faster in cold-stressed mice. Various types of cancer cells injected under the skin of mice (of the same strains used to derive the cancer cells) form tumors that grow faster when the mice are kept at a standard human room temperature (22–23°C) than when they are maintained at the higher temperatures (30–31°C) that are thermoneutral (i.e., comfortable) for mice. (*B*) This difference in tumor growth rates is not observed in severe combined immunodeficient (SCID) or nude mice, which lack a functional immune system, suggesting that the difference seen in (A) results from an effect of cold stress on the immune system's ability to fight cancer (see chapter 5). Adapted from Kokolus et al. (2013).

Because biologists routinely modify their research organism, it is not surprising that those organisms are sometimes viewed as tools, fashioned to facilitate experiments. Philosophers, too, sometimes refer to material models as "instruments for investigation" (Morrison & Morgan, 1999). It is important to remember, however, that the cells and animals used in biomedical research cannot be built entirely to experimenter specifications. They come with natural baggage that may interfere with the experiments. As noted in the next section and throughout the book, models vary in the amount of this natural baggage, and this is (and should be) a major factor in model choice.

2.3 ASSUMPTIONS ABOUT MODEL FIDELITY

Many biologists may not have an explicit philosophy of science, but they all make assumptions about their research enterprise. Particularly relevant to the use of material models in biology are assumptions about how the similarities between a model and its target vary with phylogeny and across levels of biological organization.

2.3.1 Similarity versus Phylogenetic Distance

It is widely assumed that the difference between any two species increases steadily with phylogenetic distance, such that close relatives tend to be more similar than distant relatives. Indeed, randomly evolving traits tend to diverge as lineages branch, and in aggregate the degree of similarity between any two species decreases with the square root of the phylogenetic distance between those species, defined as the time since their evolutionary divergence (Letten & Cornwell, 2014). However, this square root rule holds only when the traits evolve randomly and when the degree of similarity is averaged over many traits. Once natural selection is added to the mix, individual traits may evolve faster or slower than expected under random evolution, which then disrupts the correlation between similarity and phylogenetic distance (Striedter, 2019).

In addition, species with short generation times tend to diverge more rapidly than species that reproduce more slowly. This is important because most of the model organisms used by biologists reproduce quite rapidly and, therefore, diverge from other species (including humans) more quickly than one would otherwise expect (Bolker, 1995; J. A. Thomas et al., 2010). The relatively short generation time of rats and mice also explains why the primate genomes, overall, are more similar to those of cats, dogs, pigs, and cows than they are to rodent genomes, even though rodents are more closely related to primates than those other species (i.e., carnivores and artiodactyls) (figure 2.3). Given that evolutionary change tends to accumulate across generations, it makes sense that the rapidly reproducing rodents "evolved away" (i.e., diverged) from their nonrodent relatives so fast and so much that they are now less similar to us than our more distant, less prolific relatives.

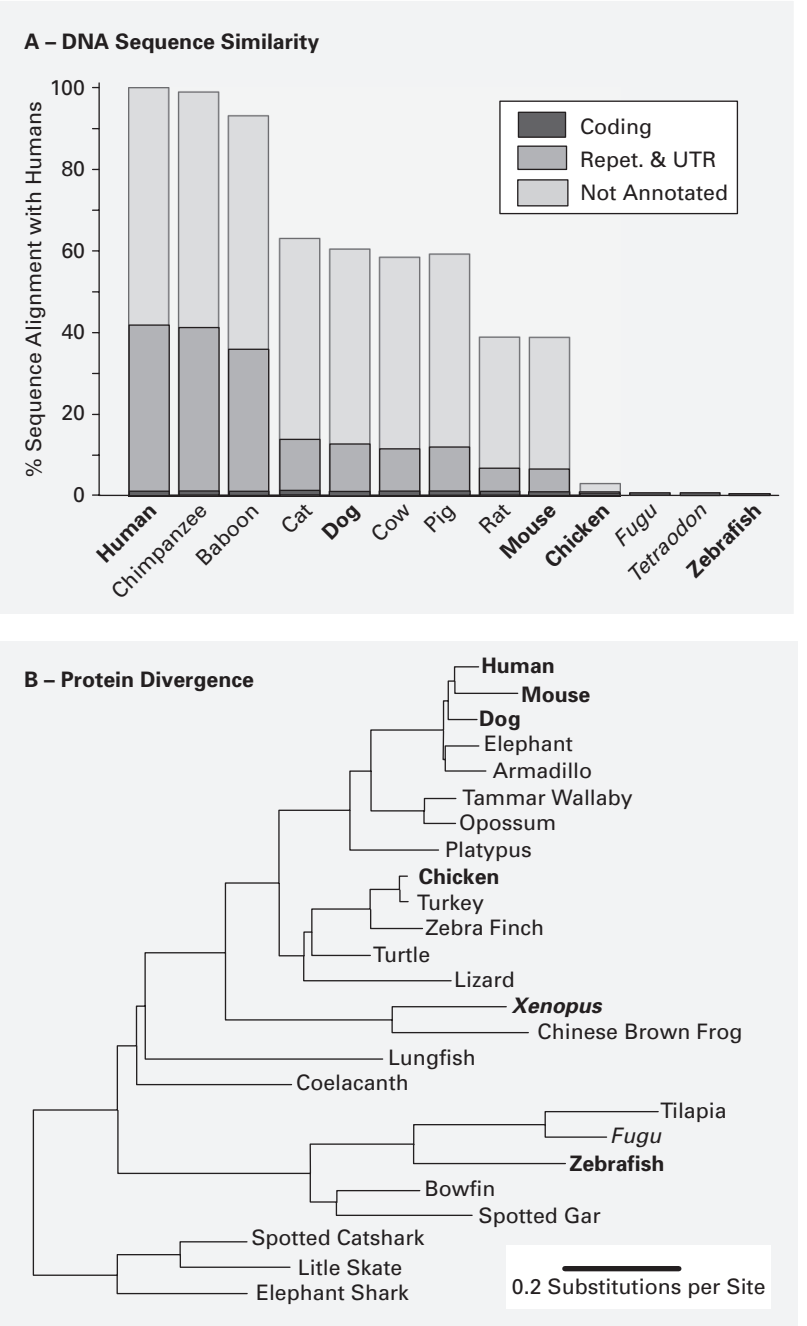


Figure 2.3

Genetic similarity versus phylogeny. (*A*) DNA sequence similarity. This histogram depicts the overall similarity of the human genome to that of various other vertebrates, using percent sequence alignment as a measure. By this measure the large animal models are more similar to humans than are rodents. UTR, untranslated region; Repet, repetitive DNA sequences. (*B*) Protein divergence. The topology of this dendrogram represents the phylogenetic relationships of various vertebrate species. The degree of protein sequence divergence (based on 243 conserved proteins) is represented by the lengths of the horizontal line segments. The dog proteins are more similar to their human counterparts than the mouse proteins, even though mice are more closely related to humans. Adapted from (A) Thomas et al. (2003); (B) Braasch et al. (2016).

Despite these complications, it still seems reasonable to believe that human diseases should, in the absence of ethical considerations (which we will discuss), be modeled in animals closely related to us, notably nonhuman primates or other mammals. Russell and Burch (of 3R fame; see chapter 1) referred to this view as the "high-fidelity fallacy" (1959) and argued, instead, that mammals do *not* always make better models for human diseases than non-mammals. Unfortunately, the examples they used to illustrate their argument were poor. A better way to address this question is to ask whether rodents are always the best nonprimate models for human diseases.

One might think that rodents are ideal nonprimate models for humans because rodents, together with rabbits and hares, comprise the mammalian lineage most closely related to primates. However, as I just noted, rodents have genetically diverged from primates more than some of our more distant relatives (figure 2.3). Furthermore, rodents tend to be much smaller than humans, and body size correlates with many therapy-related traits. It is well established, for example, that drug doses scale more reliably with body weight than phylogenetic relationship, at least among mammals (see chapter 7). Similarly, dogs and pigs are frequently used as models for cardiovascular research not because they are very closely related to humans, but because their hearts and blood vessels are very similar to ours in size (Lee et al., 2017); performing analogous experiments in mice would be extremely difficult. The existence of such size-dependent similarities between humans and dogs or pigs probably explains why the latter are often considered "higher vertebrates," even though the supposedly lower rodents are more closely related to humans.

Thus, we cannot offer a blanket statement on how similarities vary with phylogenetic relationship. It is generally true that similarities decrease with phylogenetic distance, but there are exceptions to the rule. Some of these exceptions produce greater than expected similarities (i.e., convergent similarities); others yield greater differences (e.g., in lineages with short generation times or in the form of special adaptations). Most biologists have learned to live with such exception-riddled rules.

2.3.2 Biological Hierarchies and "Disease Genes"

Despite the many obvious differences between species at the level of behavior and morphology, molecules and their interactions are often said to be highly conserved across broad swaths of phylogeny. One of the first scientists to articulate this conviction was the biochemist Alvert Jan Kluyver, who proclaimed that "from the elephant to butyric acid bacterium—it is all the same" (Friedmann, 2004). Or, as rephrased by the Nobel Prize–winning biochemist and molecular biologist Jacques Monod, "Anything that is true of *E. coli* must be true of elephants—only more so" (quoted in Friedmann, 2004). These statements were made during the early days of molecular biology but were

reinforced by many subsequent discoveries of homologous genes in widely divergent species. To quote a more recent molecular biologist, "At the outset, no one could assume that vertebrates shared so much of their molecular biology with yeast, worms and fruit flies. However, the universality of the genes and circuits that govern developmental and regulatory phenomena has since become an article of faith" (Davis, 2004, p. 73). In contrast, higher level features, especially those relating to morphology and behavior, are thought to be less broadly conserved. As Hans and John Krebs wrote in an influential paper, "it appears that at the molecular level generalizations are often valid across a very wide range of species. But at higher and more complex levels such as those of ecology and behaviour (where specialised functional adaptations have evolved) generalisations must necessarily be more restricted" (1980, p. 380).

This widespread belief in molecular conservation is evident in the notion of "deep homology" (Shubin et al., 1997). The term was coined to indicate instances in which higher level characters (e.g., behavioral or morphological features) that were not considered to be evolutionarily conserved—that is, not homologous to one another—are discovered to involve the action of homologous genes or proteins. A widely cited example is the discovery that the eyes of insects and vertebrates, which are not homologous as eyes, involve the action of *pax-6*, a broadly conserved gene (Wagner, 2007). Such findings are certainly interesting, but deep homology is sometimes interpreted as meaning that the higher level characters *must* be homologous (as higher level characters) simply *because* they involve homologous genes (Strausfeld & Hirth, 2013). Such extensions of the deep homology concept are dubious, because the deep homology might simply represent independent co-option of homologous genes into the causal networks underlying nonhomologous higher level traits (Shubin et al., 2009; Scotland, 2010). After all, it is well established that genes and proteins may change their functions during phylogeny (e.g., Liao & Zhang, 2008). In short, evolution may proceed independently at different levels of biological organization, and the homology of genes need not imply that all their functions are conserved (Striedter, 1998, 2019; Striedter & Northcutt, 1991).

A more specific and medically relevant example of scientists assuming that evolutionary conservation is more extensive at the molecular level is the notion that homologs of human "disease genes" can be studied in distantly related species and still provide important insights into the corresponding human diseases, even if the examined animals do not themselves fall ill with the disease. Thus, Rubin et al. (2000) reported that fruit flies possess homologs to 177 of 289 human disease genes, which they defined as genes that are "mutated, altered, amplified, or deleted in a diverse set of human diseases." More recent studies have reported that 80% of the human disease genes have homologs in sea anemones, and that 52% of them originated prior to the

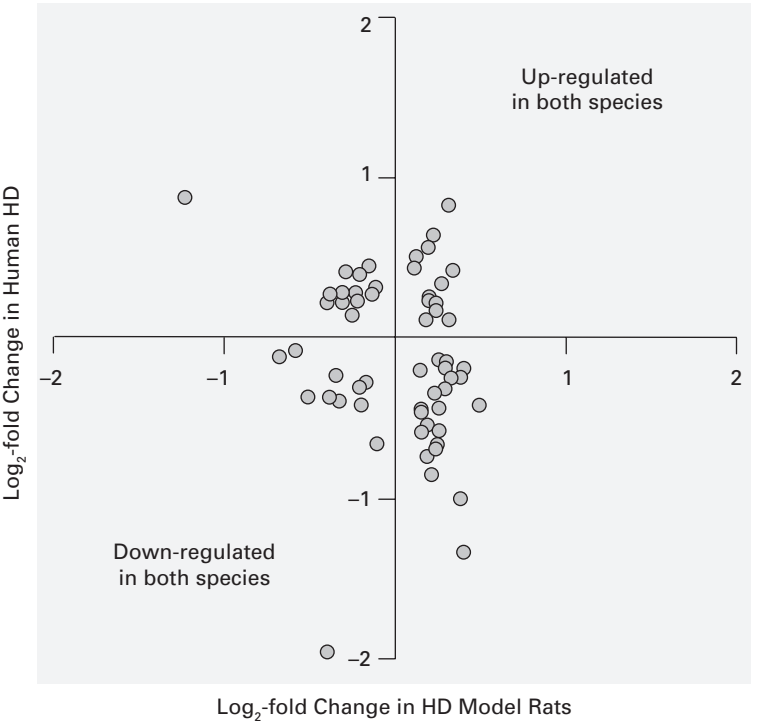
origin of animals (Sullivan & Finnerty, 2007; Maxwell et al., 2014). The broad conservation of these disease genes is often taken to imply that important discoveries about the mechanisms underlying human diseases can be made in fruit flies, sea anemones, and even more distant relatives of *Homo sapiens*, but the extent to which this assumption is warranted remains very unclear.

By contrast, it is clear that homologous disease genes and proteins have often diverged considerably in terms of their molecular structure. Most of the proteins associated with Alzheimer's disease, for example, have homologous counterparts in mice, but roughly half of them are less than 80% identical to the human proteins in terms of their amino acid sequences (Hasselmann & Blurton-Jones, 2020). In addition, molecular nonhomology is more common than one might think. For example, only 20% of all the genes in 61 different strains of the *Escherichia coli* bacterium are conserved across all of the strains (Lukjancenko et al., 2010). Nor is it simple to determine whether homologous genes and proteins are more similar than homologous morphological traits and behaviors; homologous molecules are easy to compare in terms of their percent sequence identity, but analogous comparisons for higher level traits are much more difficult and controversial. Yet another problem is that molecular functions clearly can change, not just across stages of development (e.g., Wang et al., 2014) but also across phylogeny.

The latter problem is nicely illustrated by Huntington's disease, which is a neurodegenerative disease caused by a dominant genetic mutation in a gene called *huntingtin* (*htt*). The disease appears to occur naturally only in humans, but homologs of *htt* have been identified in other mammals and several invertebrates, including fruit flies and sea urchins (Rubinsztein et al., 1994; Djian et al., 1996; Tartari et al., 2008). Research on mouse models revealed that mutant htt protein (mHtt) tends to form intracellular aggregates, which were subsequently discovered in human brains as well (Davies et al., 1997; DiFiglia et al., 1997). Experiments in fruit flies revealed additional information about how mHtt is processed inside fly cells, and how these post-translational modifications affect the protein's toxicity (Barbaro et al., 2015).

Thus, scientists have certainly learned useful information about mHtt and its downstream effects from animal models (Marsh et al., 2012). However, those effects vary considerably across species (Yu-Taeger et al., 2017) (figure 2.4). Indeed, such differences are to be expected because the *htt* gene sequences have diverged considerably over the course of evolution. Human and mouse *htt* are only about 90% identical (Barnes et al., 1994), and the fly gene has fewer than half of the exons found in human *htt* (Li et al., 1999). Moreover, both the normal and the mutant forms of Htt interact with many other proteins and genes (Langfelder et al., 2016) that themselves diverged across phylogeny. These differences help to explain why most of the potential therapies that have emerged from animal research on Huntington's disease have not borne fruit

A – Gene Expression in HD Patients vs. Model Rats



B – HD Patients vs. HD Model Rats vs. HD Model Mice

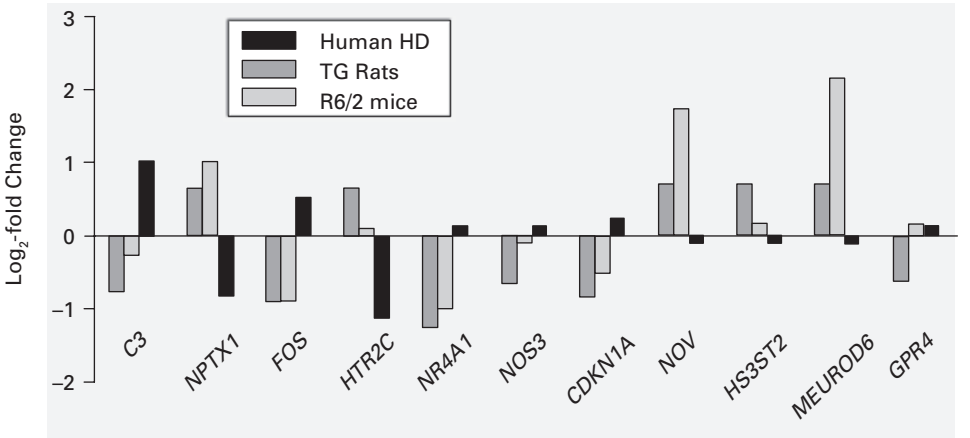


Figure 2.4

Differential gene expression in Huntington's disease (HD) patients and animal HD models. (*A*) Gene expression data from 61 genes that are up- or down-regulated in the striatum of human patients (versus human controls) and in transgenic rats expressing mutant human *huntingtin* (versus control rats). The data show that 61% of these genes were differentially expressed in opposite directions (i.e., discordantly) in the two species. (*B*) HD patients versus HD model rats versus HD model mice. This histogram compares the direction and magnitude of differential expression for 11 genes that are discordant between humans and rats (as shown in A) and also differentially expressed in HD model mice (of the R6/2 line). The mouse data consistently align with the rat data rather than the human data, as one would expect given the close relationship between the two rodent species. Adapted from Yu-Taeger et al. (2017).

in clinical trials (Wild, 2016), and why the currently most promising trials are aimed at suppressing mutant huntingtin, rather than modulating its downstream effects (for more on Huntington's disease, see chapter 6).

The important open question is, to what extent does the study of homologous disease-linked genes and proteins in diverse model systems allow for reliable extrapolations to humans? To answer this question, we must consider that disease-linked genes and proteins may change across phylogeny in terms of how they interact with other molecules, which themselves may have diverged. Such changes in gene and protein networks are likely to accumulate across phylogeny, just like evolutionary changes at other levels of biological organization. Therefore, the utility of disease models is likely to increase, at least on average, with their phylogenetic proximity to humans. However, as noted earlier, the genome as a whole has diverged at a significantly faster rate in mice than in some of the larger carnivores and artiodactyls (see figure 2.3). Thus, if the goal is human translation, it would—on average—be better to study human disease genes in nonhuman primates or, at least, other large mammals than in mice or more distant, non-mammalian relatives. Of course, as discussed later, this conclusion must be balanced against ethical and practical concerns.

2.4 DEALING WITH BIOLOGICAL COMPLEXITY

As research has progressed, the number of genes linked to human diseases has expanded to over 4,000 (Online Mendelian Inheritance in Man, 2020). Some diseases result from mutations in specific genes and run in families, but these Mendelian disorders tend to be quite rare, collectively affecting less than 5% of the population (Brunham & Hayden, 2013). The more common diseases, by contrast, tend to be associated with a large number of genes that, individually, contribute relatively little to disease risk and may be swamped in their effect by nongenetic factors such as smoking and nutrition (Joyner, 2011). Therefore, even diseases that have a substantial genetic component often have relatively common "sporadic" forms that cannot be predicted from family histories. Indeed, most minor risk genes were identified only through very large genome-wide association studies that examined thousands of patients.

In general, it remains quite difficult to specify how individual genes contribute to their associated diseases. Even in Huntington's disease, where mutation of a single gene causes the disease, it remains a serious challenge to explain how this mutation leads to neurodegeneration and, ultimately, death. The reductionism of molecular biology excels at identifying disease-related genes and proteins, as well as their potential molecular interactions, but explaining how you get from the molecular level to the

level of organismal physiology and disease progression is more difficult. In the words of Salvador Luria, a founding father of molecular biology, "Everyone keeps going down, down, down, trying to be more reductionist, trying to see finer and finer, to find the basis of structure and function. . . . I think it's time start going up again—going in the opposite direction" (quoted in Weiner, 1999, p. 213). This is easier said than done, of course. As Seymour Benzer, who became famous for his work on the genetic mechanisms underlying learning and memory in flies, reportedly once said, "It takes a long time of going down before you start looking to go up again. Down is a much easier way to go" (quoted in Weiner, 1999, p. 213).

2.4.1 Endophenotypes

In an effort to manage this problematic aspect of traditional genetic reductionism, some biologists have proposed a middle-out approach that starts at some middle level of the biological hierarchy and then extends downward to the level of genes as well as upward to higher biological levels. Features at this middle level are called endophenotypes because they "occupy the terrain between disease symptoms and risk genotypes" (Insel & Cuthbert, 2009). An explicit focus on these endophenotypes was recently endorsed by the National Institute of Mental Health for the study of mental disorders (Kozak & Cuthbert, 2016). Part of the motivation for this move was the recognition that the diagnostic criteria for most psychiatric disorders (aka mental or behavioral disorders) do not align in any simple way with the findings emerging from genetics and neuroscience, thereby impeding the classic reductionist approach. Another motivator was that "animals will never have guilty ruminations, suicidal thoughts, or rapid speech. Thus, animal models based on endophenotypes that represent evolutionarily selected and quantifiable traits may better lend themselves to investigation of psychiatric phenomena than models based on face-valid diagnostic phenotypes" (Gottesman & Gould, 2003, p. 641).

Schizophrenia, for example, is essentially impossible to model fully in nonhumans because so many of its symptoms involve higher mental processes (see chapter 6). Therefore, researchers have focused on creating animal models of simpler processes that are thought to be components of human schizophrenia, such as deficits in sensorimotor gating, learning and memory, as well as anxiety and sleep disturbances (Tarantino, 2000). The hope is that the animal studies will elucidate the mechanisms underlying these endophenotypes and that the aggregate of all this information will explain human schizophrenia and lead to more effective treatments. Moreover, researchers hope that some of the same endophenotypes will also contribute to other complex mental disorders, such as autism spectrum disorder. Whether these aspirations will be fulfilled remains unclear. One may doubt, for example, that complex mental disorders are caused by

linear combinations of endophenotypes; it seems more likely that some endophenotypes, like many of their underlying genes, can interact in a nonadditive manner.

An interesting twist on the endophenotype approach is provided by research on narcolepsy, a disease that is defined by chronic drowsiness and sudden attacks of sleep. Several breeds of dogs suffer from a very similar, likely homologous disease, and those dogs all have mutations in the gene for a receptor of a signaling peptide called hypocretin (Lin et al., 1999). Subsequent research showed that deletion of the hypocretin (aka orexin) gene in mice correlates with increased sleepiness (Chemelli et al., 1999). Importantly, later studies showed that humans with narcolepsy tend to lack hypocretin-expressing neurons (Thannickal et al., 2000). Thus, the animal research in this case made an accurate prediction about an endophenotype underlying a human disease.

The thought-provoking twist is that human narcolepsy patients generally do not exhibit mutations in the hypocretin or hypocretin receptor genes (Bućan & Abel, 2002). Indeed, the human disease is generally not what geneticists call familial (i.e., inherited), and considerable evidence suggests that the major type of narcolepsy in humans results from immune cells attacking the body's own hypocretin neurons (Kornum & Jennum, 2020). In short, the animal models and human narcolepsy patients share the disease and a disease-causing endophenotype, but they differ in the endophenotype's causal basis; apparently there are multiple ways to kill the hypocretin-expressing neurons or render them dysfunctional.

2.4.2 Systems Biology

Biological research has also been transformed by the rise of systems biology, which may be defined as "studying the interrelationships of all of the elements in a system rather than studying them one at a time" (Hood, 2003, p. 9). A common aim of systems biology is to integrate vast amounts of data at various levels of biological organization, especially at the molecular levels, by using statistical methods to identify pathways and modules of coregulated or interacting molecules (e.g., Langfelder et al., 2016; Ament et al., 2017, 2018). These molecular network features are then linked to higher level functions through correlational analyses, targeted manipulations, and gene ontology databases that collate information about the functions of individual genes. Some aspects of this systems biology paradigm have been criticized as "low-input, high-throughput, no-output science" (see Friedberg, 2008). However, systems biology is changing rapidly and promises to yield useful output by identifying critical modules, hubs, or choke points at which a system can be manipulated most effectively for therapeutic purposes (Langfelder et al., 2013; Swarup et al., 2019).

The basic tools of systems biology have been employed for more than a decade in the field of environmental toxicology (National Research Council, 2007). During

this time, the field largely moved away from testing potentially toxic compounds in animals toward examining their effects on cultured cells, preferably of human origin. In particular, researchers examined whether those chemicals activated molecular toxicity pathways, also known as "adverse outcome pathways" (Jeong & Choi, 2018). A major benefit of this widely heralded paradigm shift was that it greatly reduced animal suffering and increased the rate at which compounds could be tested. The major drawback thus far has been a paucity of studies examining how well the various in vitro toxicity assays predict toxicity in animal models or humans. As a result, confidence in the new approaches appears to remain relatively low (Ginsberg et al., 2019). Indeed, regulatory decisions still tend to be based on both in vitro and animal tests as well as human epidemiological data. Despite these challenges, there have been calls to extend the adverse outcome pathways approach from toxicology to disease-oriented research (Langley et al., 2015). At that point, this approach merges with general systems biology, although the latter is less committed to the replacement of animal research with in vitro and in silico studies.

Another promising direction for systems biology is to develop mathematical models of complex molecular systems. The mathematical models allow investigators to make predictions about how the target system should respond to perturbations (Hood, 2003). The mathematical models themselves can also be examined for unexpected properties. A particularly interesting finding to emerge from such studies is that the models of many biological systems tend to be insensitive to variations in most of their parameters, a phenomenon referred to as "sloppy control" (Gutenkunst et al., 2007). This discovery implies that most systems can be controlled by varying a relatively small set of key parameters. Moreover, a system's sensitivity to changes in any one parameter depends to a surprisingly large degree on the values of the system's other parameters (Wagner, 2015), which is to say that the importance of any given parameter is highly context dependent. Finally, simulated evolutionary changes in a model's parameters can quickly change the distribution of parameter sensitivities, such that previously very important parameters become unimportant. This observation might explain why genes that are essential for survival or reproduction in one species or cell line are often nonessential in others (Liao & Zhang, 2008). More generally, these findings reinforce the notion that extrapolations of gene functions across species should be made cautiously (see chapter 7).

2.5 ANIMAL WELFARE AND COGNITIVE DISSONANCE

Experimental biologists are "often portrayed by animal rights groups as having little or no regard for animal welfare" (Franco et al., 2018, p. 14), but most biologists do care

about the welfare of their research animals. Not all of them are as eager as the environmental toxicologists we just discussed to replace animal models with nonanimal alternatives, but most accept that doing so is a good long-term goal (Franco et al., 2018). In several countries (especially Japan) researchers even hold memorial ceremonies for their research animals (Narver et al., 2017). In that sense, biology has come a long way.

2.5.1 From "Can They Think?" to "Can They Suffer?"

The Romans killed thousands of animals in their circuses, including bears, tigers, and elephants, forcing them to fight with one another or with gladiators (Singer, 1990). The early Christians showed occasional flickers of concern for animals, but St. Thomas Aquinas represented the traditional Christian position when he proclaimed that "it matters not how man behaves to animals, because God has subjected all things to man's power . . . and it is in this sense that the Apostle says that God has no care for oxen" (1947, I-II, Q102, A6, ad.8). Sustained concern for animal welfare and rights did not emerge in Europe until the 18th century (French, 1975), though it arose earlier in other parts of the world with other religious traditions (Caruana, 2020). Particularly important was the contribution of the philosopher Jeremy Bentham. Before him, most Western philosophers had thought that animals deserved much less consideration than humans because, as Plato had argued, only humans have a "rational soul" (Rosenfield, 1940; Cottingham, 1978; Smith, 2010). Bentham changed this calculus by arguing that "the question is not, Can they *reason?* nor, Can they *talk?* but, Can they *suffer?*" (1789, p. 311; emphasis in original). He concluded that it would be acceptable to slaughter animals rapidly and relatively painlessly, at least for food, but that there was no justification for, as he put it, tormenting them. But which animals are capable of suffering? More generally, which ones have feelings? To use more technical language, which ones are sentient?

Answers to these questions cannot be obtained by direct observation of animal behavior because expressions presumed to be of pain or joy are not necessarily associated with conscious feelings (Hatfield, 2007). Language can give us some insight into the feelings of other humans, but without linguistic communication, the "other minds problem" becomes a Gordian knot (Avramides, 2019). Many people are willing to give other mammals and birds the "benefit of the doubt" (Harnad, 2016) with regard to sentience, perhaps because those animals are warm-blooded and cute, especially when they are young. But what about cold-blooded reptiles, fishes, or invertebrates? Does sentience come in degrees that correlate with brain size or behavioral complexity, as some have suggested (de Waal, 2019)? I tend to think so, but this idea is difficult (if not impossible) to demonstrate rigorously. In general, our willingness to attribute sentience to animals seems to fall off with their phylogenetic distance from us, but this is not exactly true—birds are more closely related to cold-blooded reptiles than to

mammals. It is more accurate to state that most people attribute sentience in accordance with where they place animals along a vaguely defined continuum commonly known as the "phylogenetic scale." The obvious problem with this idea is that the notion of a phylogenetic scale has long been debunked (Hodos & Campbell, 1969).

Evolution produces bushes rather than ladders, and different ranking criteria lead to different arrangements of species along putative scales. To illustrate, consider cephalopods (octopuses and squids): as invertebrate mollusks, they are only distantly related to vertebrates, yet they have remarkably complex nervous systems and behavior. So where do they fit on the supposed phylogenetic scale? Recognizing this nonlinearity problem, the European Union recently labeled cephalopods "an exceptional invertebrate class" and gave them special research protections (Berry et al., 2015).

Another problem with the phylogenetic scale is that most humans rank pests much lower than pets, irrespective of their phylogenetic position. Most glaringly, most humans are far more inclined to protect cats and dogs than mice or rats, even though cats and dogs are more distantly related to humans (Foley et al., 2016). Perhaps we should replace the phylogenetic scale with a genetic divergence scale in which species are ordered by their overall genetic similarity to humans; according to such a scheme, dogs and pigs would rank higher than mice and rats (see figure 2.3). However, such a genetic divergence scale has problems of its own. Moreover, many genes and other traits clearly do follow phylogeny, for those "phylogenetic signals" are what scientists use to reconstruct phylogenies.

2.5.2 Animal Welfare Laws and Regulations

Given the complicated and evolving nature of our collective views on animal sentience, it is not surprising that the laws and regulations concerning animal experiments have changed over the years and are not completely rational. For example, the first significant animal welfare legislation, the United Kingdom's Cruelty to Animal's Act of 1876, extended protections to all vertebrates, but its initial draft had excluded all cold-blooded animals, including reptiles, fishes, and frogs. By the time the act was passed, it applied to all vertebrates and excluded all invertebrates. This changed in 1993, when the United Kingdom decided to regulate experiments on the common octopus; those regulations were extended to all cephalopods in 2013 and, as mentioned in the preceding section, now apply throughout most of Europe.

The situation in the United States is more complex. The current version of the Animal Welfare Act (US Department of Agriculture, 2020) protects all warm-blooded animals but includes the following proviso:

> The term "*animal*" means any live or dead dog, cat, monkey (nonhuman primate mammal), guinea pig, hamster, rabbit, or such other warm-blooded animal, as the Secretary [of the USDA] may determine is being used, or is intended for use, for research, testing,

> experimentation, or exhibition purposes, or as a pet; but such term excludes *(1)* birds, rats of the genus Rattus, and mice of the genus Mus, bred for use in research, *(2)* horses not used for research purposes, and *(3)* other farm animals, such as, but not limited to livestock or poultry, used or intended for use as food or fiber . . . (7 U.S.C. §2132(g))

In response to public pressure, the US Congress in 1970 gave the US Department of Agriculture (USDA) the authority to include mice and rats in its regulations, but the USDA declined. Because the USDA collects animal numbers only for the species under its purview, it is difficult to obtain precise estimates of the total numbers of animals used for research in the United States. The best one can currently do is to request records from individual research institutions about how many animals of the various species they are housing on any given day, on average (figure 2.5A). By comparison, the United Kingdom publishes annual reports on how many procedures were approved for the various vertebrate species, plus cephalopods (figure 2.5B).

Despite the species gaps in record keeping by the USDA, the use of all vertebrates in biological research in the United States is regulated by the Public Health Service, which administers the National Institutes of Health (NIH) and thus provides most of the support for US animal research. As a result, all research institutions that receive money from the NIH must follow its regulations regarding animal research. Still, mice and rats continue to be exempted from federal inspections and reporting requirements.

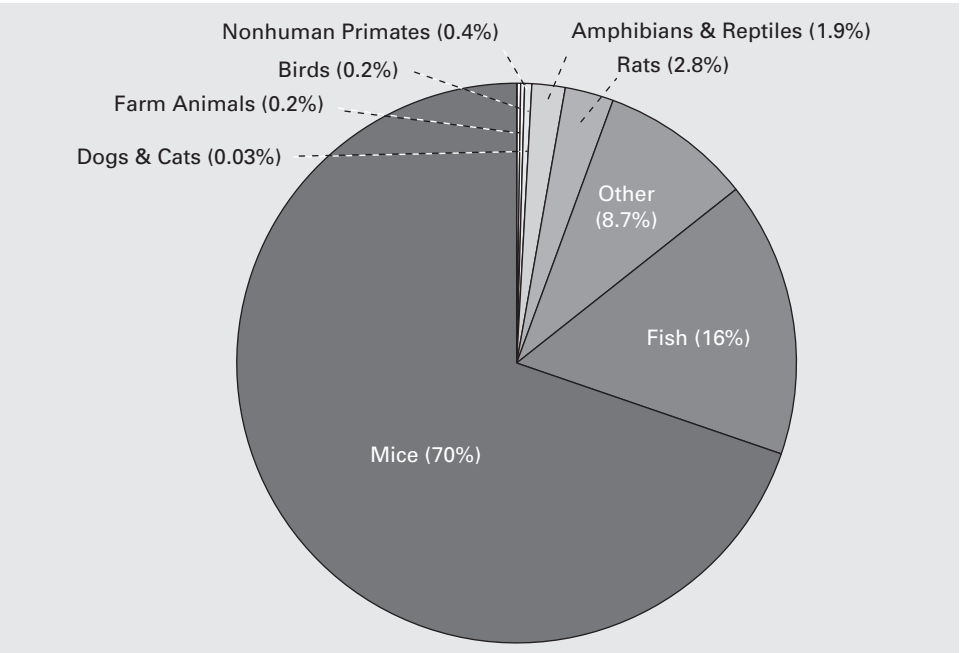
2.5.3 Conflicting Attitudes

Ethical concerns frequently conflict with practical concerns, especially for experimental biologists, farmers, and people who eat meat or need medicines derived from animal research. Even Mary Beth Sweetland, a former vice president of People for the Ethical Treatment of Animals (PETA), regularly took insulin for her diabetes, even though the discovery of insulin was clearly based on animal experiments. She may have justified this choice by arguing that she needs her life to fight for the rights of animals, but the case still illustrates the internal conflicts that animal experiments may spawn. Importantly, many people harbor such conflicts without being fully aware of them, creating cognitive dissonance.

In general, animal rights advocates tend to believe that animals and humans have roughly the same capacity for suffering and thought, but many also claim that animals are so different from humans that experiments on animals can tell us nothing about human biology. These beliefs are inconsistent with one another, unless we are prepared also to believe that feelings and thought are based on something other than biological processes.

Biologists, in contrast, tend to have the opposite problem (Shapiro, 2004). They often think that animals and humans are so similar to one another that the former can

A – Animal Inventory in US Research Institutions



B – UK Experimental Procedures by Species, 2018

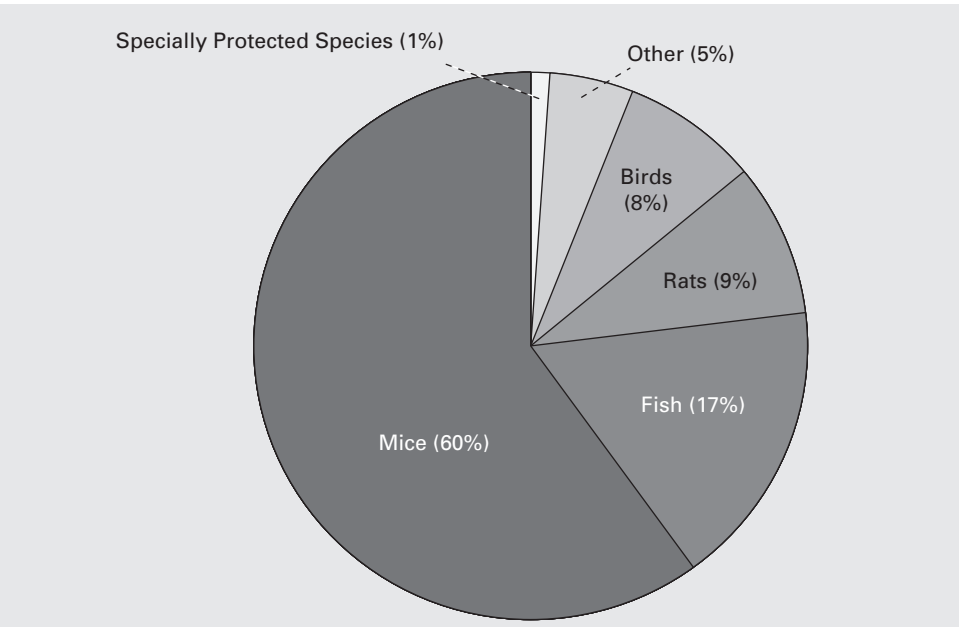


Figure 2.5

Frequency distribution of vertebrates used in research. (*A*) Animal inventory in US research institutions. Because the US government does not track the number of rodents and birds in research, Goodman et al. (2015) obtained average daily animal inventory numbers from 21 major US research institutions from 2008–2012. Mice clearly dominate this data set, which is limited to vertebrates. (*B*) Distribution of vertebrate species used for experimental procedures in the United Kingdom in 2018, as reported by the Home Office. The specially protected species include nonhuman primates, dogs, cats, and horses. Based on data in (A) Goodman et al. (2015); (B) Home Office (2019).

serve as models for the latter in most instances. Yet biologists also tend to believe that the suffering of a mouse, for example, is not as ethically troubling as that of a human. The resulting dissonance was clearly recognized by Anna Kingsford, one of the first women ever to obtain a medical degree, as well as an influential vegetarian and anti-vivisectionist. As she wrote in 1883,

> it is precisely the subtle but enormous differences existing between the manifestations and character of the nervous system as we see them in man and as we see them in other animals, which distinguishes the former from the latter, and which endows vivisectors with the legal right they now possess to inflict on anthropoid apes injuries and mutilations which, if they inflicted the same on men, would be held to render the perpetrators guilty of crime. When, therefore, it is understood that this occult nervous differentiation is capable of constituting a distinction so vast, how is it possible to suppose that the study of biological function in the beast is capable of explaining satisfactorily the mysteries of human life? (cited in French, 1975, p. 316)

Compared with the concerns expressed over animal experiments, sympathy for animals we use for food is rather limited, even though farm animals outnumber laboratory animals by at least two orders of magnitude. Despite major gaps in the reporting of animal numbers, the number of vertebrate animals used in research worldwide is probably below 100 million per year (Statista Research Department, 2016). By comparison, humans consume approximately 50 billion chickens per year, excluding all the male chicks that are killed right after hatching because they cannot lay eggs and are not needed for breeding (Thornton, 2019).

Importantly, virtually all these chickens are killed without anesthesia, which is not the practice in animal research. Particularly brutal is the mass killing of animals when herds or flocks get infected with serious diseases. For example, during the 2003 outbreak of Newcastle disease, the State of California ordered 3.1 million chickens to be killed within a few weeks. One farmer killed 30,000 of his chickens in a wood chipper, an act that was later judged to be "standard industry practice" (Gillick, 2003). Similarly, farmers in just one US state (Minnesota) are estimated to have killed 90,000 excess hogs during six weeks in early 2020 because local meatpacking plants were closed on account of the COVID-19 pandemic (Bailey, 2020); by comparison, US biologists kill roughly 50,000 pigs for research every year (APHIS Annual Report Animal Usage by Year, 2019).

A cognitive dissonance that is of more direct and practical concern to experimental biologists is the push from federally mandated institutional animal care and use committees (IACUCs) to *reduce* the number of animals used in research, while statisticians are telling the biologists that their studies are underpowered and should include

more animals, including both sexes (see chapter 1). Similarly, IACUCs are instructing biologists not to duplicate existing research, while the ongoing replicability crisis has highlighted the need for more replication. In short, biologists are regularly confronted with a variety of conflicting exhortations.

2.6 COMPARATIVE BIOLOGY AND AUGUST KROGH REVISITED

Although popular views and regulations concerning animal welfare are imbued with the idea of a phylogenetic scale, which is centered on species differences, many biologists are more engrossed with species similarities. Especially molecular biologists have traditionally been interested only in "universals" and considered variation to be uninteresting noise (Orzack & McLoone, 2019). As Rowland Davis (2003) put it in his book on the use of model microbes in biology,

> The rise of model organisms had an ironic effect. The reductionist approach, in favoring simple systems, brought scientists to model microbes, and genetic rationales trapped its professionals in their use. The result was a concentration on a few organisms explicitly used to discover generalizable principles. Indeed, geneticists, in standardizing stocks and media, provided artificial representatives of various species living in an artificial environment. . . . the appreciation of diversity and complexity was lost for some time among molecular biologists. A curious construction of biology arose in which many biologists spoke of "the cell" or "the organism" when in fact they were studying *N. crassa* or *E. coli* or *S. cerevisiae.* A Platonic view of life emerged in which model organisms were the reality and the rest of the living world was a chaos of variants and exceptions. The irony is that even as neo-Darwinian views of evolution freed us from such thinking, molecular biology acquired an even more rigorous typological stance than that of Linnaean taxonomists of the early 19th century. (p. 254)

Modern molecular biology has broadened its focus, but a strong emphasis on similarities continues to pervade much of biology. It is even inherent in Krogh's principle (see section 2.2) because John Krebs, who named this principle, expected the discoveries made in the "most convenient" experimental species to generalize across species, at least at the molecular and cellular levels (Krebs & Krebs, 1980). This is why he considered the study studied species to be "examples" rather than "models" (Krebs, 1975). It is true that "Krogh organisms" (Green et al., 2018) are typically selected for having extreme adaptations, which is what makes them so accessible to inquiry, and that the study of these creatures "may reveal general principles not readily observable in less extreme species" (Pollak, 2014, p. 442). However, the discovery of general principles does not preclude the existence of fundamental species differences. To underscore this point, consider Krogh's own research.

August Krogh was a comparative physiologist who studied a wide variety of organisms, including many invertebrates, and focused much of his research on the topic of gas exchange (Krogh, 1941). How do animals absorb oxygen and shed carbon dioxide? How do the mechanisms they adopt vary with the environment in which they live (e.g., aquatic versus terrestrial habitats), with body size, and with specific aspects of their anatomy? To answer such questions, Krogh used general principles, such as those governing gas diffusion and absorption, as well as scaling laws that govern surface-to-volume ratios. His principal goal was not to discover new universals but to reveal how such principles were implemented differently in different species. He wanted to know what solutions evolution brought forth, given the constraints of physicochemical laws and biological structure-function principles. Given this background, it becomes clear why Krogh wrote that "the route by which we can strive toward the ideal [state of general physiology] is by a study of the vital functions in all their aspects throughout the myriads of organisms" (1929, p. 202). He sought to explain variation in terms of general principles, which is quite different from studying the most convenient species to understand them all.

In short, Krogh advocated for a truly comparative approach that focuses on differences as much as similarities. Physiology has generally retained this orientation, but other fields have not. For example, Daniel Lehrman lamented the loss of a truly comparative approach in behavioral research during the heyday of behaviorism: "The value of comparison comes not from the merging of different levels into a misleadingly unified conception of behavior but from the development of an evolutionary perspective which enables us to appreciate the emergence of new qualities without neglecting the underlying continuities and their transformations" (Lehrman, 1971, p. 468). Of course, this tension between those who seek universals and those who emphasize variation is hardly new—nor confined to biology. As Francis Bacon wrote in 1620,

> The greatest and, perhaps, radical distinction between different men's dispositions for philosophy and the sciences is this, that some are more vigorous and active in observing the differences of things, others in observing their resemblances . . . each of them readily falls into excess. (cited in Friedmann 2004)

As detailed in chapter 7, the kind of truly comparative perspective advocated by Lehrman and Krogh helps explain the current translatability crisis and suggests a more productive way forward. But first, let us explore how biologists have used diverse material models to address human diseases and other frailties.

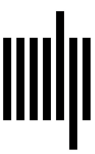
Model Systems in Biology

History, Philosophy, and Practical Concerns

By: Georg Striedter

Citation:
Model Systems in Biology: History, Philosophy, and Practical Concerns
By: Georg Striedter
DOI: 10.7551/mitpress/14366.001.0001
ISBN (electronic): 9780262370028
Publisher: The MIT Press
Published: 2022

The open access edition of this book was made possible by generous funding and support from MIT Press Direct to Open



The MIT Press

The MIT Press would like to thank the anonymous peer reviewers who provided comments on drafts of this book. The generous work of academic experts is essential for establishing the authority and quality of our publications. We acknowledge with gratitude the contributions of these otherwise uncredited readers.

This book was set in Adobe Garamond Pro and Berthold Akzidenz Grotesk by Westchester Publishing Services.

Library of Congress Cataloging-in-Publication Data

Names: Striedter, Georg F., 1962– author.
Title: Model systems in biology : history, philosophy, and practical concerns / Georg Striedter.
Description: Cambridge, Massachusetts : The MIT Press, [2022] | Includes bibliographical references and index.
Identifiers: LCCN 2021033979 | ISBN 9780262046947 (hardcover)
Subjects: LCSH: Animal models in research. | Animal experimentation.
Classification: LCC R853.A53 S77 2022 | DDC 616.02/7—dc23
LC record available at https://lccn.loc.gov/2021033979