

---

# *The Philosophical Quarterly*

---

## TWO MODELS OF MODELS IN BIOMEDICAL RESEARCH

BY HUGH LAFOLLETTE AND NIAL SHANKS

Biomedical researchers claim there is significant biomedical information about humans which can be discovered only through experiments on intact animal systems (AMA p. 2). Although epidemiological studies, computer simulations, clinical investigation, and cell and tissue cultures have become important weapons in the biomedical scientists' arsenal, these are primarily 'adjuncts to the use of animals in research' (Sigma Xi p. 76). Controlled laboratory experiments are the core of the scientific enterprise. Biomedical researchers claim these should be conducted on intact biological systems, whole animals. By observing the effects of various stimuli in non-human animals, we can form legitimate expectations about the likely effects of these stimuli in humans. Perhaps more importantly, we can understand the biomedical condition's causal mechanisms.

This reigning view of animal experimentation construes animal models as *causal analogue models* (CAMs). Yet there is another view of animal experimentation which, although not entrenched in biomedical theory, is alive and well in biomedical practice. On this view animal test subjects may not be similar to human phenomena in relevant causal respects, and thus do not *prove* or *establish* anything about human beings. Rather, experiments on animals prompt the formation of hypotheses about the nature of biomedical phenomena in humans. Such animal models we call *hypothetical analogical models* (HAMs).<sup>1</sup>

<sup>1</sup> In earlier papers (1993a, 1993b), we distinguished CAMs from what we called 'heuristic devices', adopting this term from Hempel (see the quotation in the next paragraph). However, since talk of 'heuristic devices' has taken on a particular technical usage under the influence of Imre Lakatos, we have decided our earlier usage of this term could be misleading. Hence our introduction of 'HAMs'.

Hypothetical analogical models are exceedingly valuable scientific devices, especially in the context of discovery. As Hempel remarks (p. 441):

More important, well-chosen analogues or models may prove useful 'in the context of discovery', i.e., they may provide effective heuristic guidance in the search for new explanatory principles. Thus, while an analogical model itself explains nothing, it may suggest extensions of the analogy on which it was originally based.

In this paper we explore these two types of analogical reasoning in the biomedical sciences. After sharpening the distinction between HAMs and CAMs, we critically examine the use of non-human CAMs to test hypotheses (often derived from non-human HAMs) about the causal mechanisms underlying human biomedical phenomena. We argue that although animal models are scientifically legitimate HAMs, they are probably not suitable CAMs.

## TWO MODELS OF MODELS

### 1. CAMs

The standard view of animal experimentation in biomedical research is traceable to Claude Bernard, cited by the AMA as having set down the principles guiding biomedical research, who says (p. 125):

Experiments on animals with deleterious substances or in harmful circumstances are very useful and entirely conclusive for the toxicology and hygiene of man. Investigations of medicinal or of toxic substances are also wholly applicable to man from the therapeutic point of view; for as I have shown, the effects of these substances are the same on man as on animals, save for differences in degree.

On this view the primary function of animal tests is to uncover the causal mechanisms which produce and direct the course of a disease or condition in animals. These results are then extended by analogy to humans. The resultant understanding of the relevant causal mechanisms in humans empowers scientists to prevent or treat the disease or condition under investigation. Consequently, CAMs are thought to be the primary engine of biomedical advance. According to the AMA (p. 16):

virtually every advance in medical science in the twentieth century, from antibiotics and vaccines to antidepressant drugs and organ transplants, has been achieved either directly or indirectly through the use of animals in laboratory experiments.

## 2. *HAMs*

Hypothetical analogical models play a significant role in science, especially in the early stages of a science. For instance, the planetary model of the atom, according to which the electrons in atoms orbited the nucleus much as planets orbit their sun, played a pivotal role in the early years of the study of the atom.

There is no doubt that this analogy ... was extraordinarily fruitful during the first half of the twentieth century. It suggested all sorts of questions that formed the basis of much research. For example, 'How fast are the electrons moving around in their orbits?', 'Are the orbits circular or elliptical?'. In investigating such questions, scientists learned much about atoms. In particular they learned about many respects in which atoms are not like the solar system. In the end, a good analogy often leads to its own demise (Giere p. 24).

HAMs also played important roles in molecular biology:

In *The Double Helix*, Watson talks about noticing spiral staircases, and of thinking that the structure of DNA might be like a spiral staircase. He also had the example of Pauling's  $\alpha$ -helix. Here we would say that Watson was using the spiral staircase and the  $\alpha$ -helix as analogue models for the DNA molecule .... One might also say he was modelling the structure of DNA on that of the  $\alpha$ -helix or a spiral staircase (Giere p. 23).

HAMs likewise played a role in the history of biomedicine. Elie Metchnikoff developed his cellular theory of immunity after observing digestion in the 'mobile cells of a transparent starfish larva'. Although the causal mechanisms of larvae cells are not at all akin to the causal mechanisms of phagocytic cells (the first line of defence against invading organisms), their functional similarities enabled Metchnikoff to gain new insights about the nature of human immunity (Silverstein p. 44).

### 3. *What makes a scientifically legitimate HAM?*

Scientifically legitimate HAMs are not merely psychological causes which serendipitously prompt a scientist to make a discovery. As a matter of fact, a scientist might gain important insights about the metabolism of phenol after jogging a mile, listening to Beethoven's Fifth, or drinking a cup of coffee. But that does not mean jogging, listening to music or drinking coffee is the same as studying a HAM. There is no particular reason why these activities prompted the scientific insight; nor do we have any reason to think they would prompt important insights by other

scientists. These are merely unique psychological causes, not scientific devices.

On the other hand, scientists plausibly assume that experiments on animals can suggest fertile hypotheses about biomedical phenomena. A HAM is valuable in as much as there are demonstrable functional similarities between the model and item modelled. Since there are demonstrable functional similarities between humans and our close biological relatives, biomedical scientists infer that the results of tests on animals will probably prompt ideas about how to think about and understand the functionally analogous human phenomena. For example, scientists might observe that pigs metabolize phenol primarily through glucuronidation conjugation reactions (95%) and subsequently hypothesize that humans do likewise. This plausible hypothesis, however, turns out to be false. We discover that humans metabolize only 12% of phenol in this way. Other scientists observe that rodents metabolize 45% of phenol through conjugation with sulphate and subsequently hypothesize that humans do likewise. This alternative hypothesis turns out to be closer to the truth: humans metabolize 80% of phenol by reactions with sulphate.

Although only the second hypothesis was (partially) confirmed by the data, both were reasonable predictions from the evidence to hand. For not only do we observe functional similarities between humans and other mammals, the theory of evolution suggests that phylogenetic kin (creatures descended from common ancestors) will share many of the same functional properties. That is why mammals, our phylogenetic cousins, are usually the HAMS of choice. That is why scientists assume that mammals will oxygenate the blood in roughly the same way, that they will excrete wastes in much the same way, etc.

#### 4. *Testing the hypothesis*

Basic scientists can use the results of animal experiments to propose hypotheses about functionally similar human biological phenomena. Those hypotheses must then be tested. The most direct way to test hypotheses *about* humans is to conduct tests *on* humans. And that is what researchers do in clinical and epidemiological studies. However, although there are established retrospective and prospective research methodologies for such studies, many biomedical researchers advocate using non-human models as CAMs instead. Why?

Scientists typically think experiments must be tightly controlled if they are to yield reliable data. Thus these scientists eschew research methodologies which are not as tightly controlled. Moreover, since most people

think controlled toxicological or teratological experiments on humans would be immoral, scientists turn to controlled experiments on non-human animals as the CAMs of choice.

A researcher who fails to distinguish HAMs from CAMs is unlikely to think there is anything illicit about that choice. More specifically, the researcher might assume that since humans and mammal test-subjects are phylogenetically related, then not only are they functionally similar, but moreover that the underlying biological mechanisms are likewise similar – if not nearly identical. But there is a big difference between an animal model's being a good source of hypotheses and its being a good means to test hypotheses.

### CAMS: REFINEMENTS AND DIFFICULTIES

#### *1. Extrapolating from laboratory experiments*

To evaluate the use of non-human CAMs as tests for hypotheses about human biomedical phenomena, we must ascertain the conditions under which we can plausibly extrapolate from laboratory experiments to 'real world' phenomena. When as chemists or physicists we conduct a laboratory experiment, we manipulate some substance *X* and record the results. Then, using the principles of standard causal determinism (all events have causes, and, for qualitatively identical systems, the same cause produces the same effect) we infer that similar manipulations of *X* outside the laboratory will have similar effects. For instance, we combine hydrogen with oxygen in the laboratory: water is produced. We infer that when similar elements combine outside the laboratory, water will likewise be formed. It is a sound inductive inference.

Nineteenth-century biomedical researchers imported these methodological presuppositions from the physical sciences. As Bernard expresses it (p. 148):

We cannot imagine a physicist or a chemist without his laboratory. But as for the physician, we are not yet in the habit of believing that he needs a laboratory; we think that hospitals and books should suffice. This is a mistake; clinical information no more suffices for physicians than knowledge of minerals suffices for chemists and physicists.

The inferential mechanism borrowed from the physical sciences may capture some very simple biological processes. Perhaps all cells react similarly to certain acids. However, this model cannot capture most biological phenomena since they are probabilistic. That is, an experimenter observes some phenomenon in a certain percentage of the laboratory

subjects of type  $X$  and infers that a certain percentage of creatures of type  $X$  will react similarly outside the laboratory.

There are those who think that probabilistic reasoning is bogus, that it cannot legitimately be causal reasoning since it does not fit the model of Humean 'constant conjunction'. We see no reason to embrace such a restrictive view of causality. We agree with Wesley Salmon (p. 190) that *probabilistic causality* is a 'coherent and important scientific concept', and (p. 188) that there is 'compelling (though not absolutely incontrovertible) evidence that cause-effect relations of an ineluctably statistical sort are present in our universe'. However, we do not wish to engage in a defence of Salmon's position; we shall simply assume that such a view is acceptable. In the context of the current paper, this is an innocent assumption, especially since the scientific legitimacy of most biomedical research depends on it. If the assumption were mistaken, then the overwhelming majority of biomedical experiments – whether animal or human studies – would not be justified.

Do these two options embody the methodology of animal experimentation? Researchers apparently think so. They believe inferences from non-human CAMs to humans utilize normal causal reasoning. However, animal experimentation is neither straightforwardly deterministic nor probabilistic in the senses discussed above. In both instances, experimenters make inferences from what happens to  $X$ s in the laboratory to what will happen to  $X$ s outside the laboratory. Not so with animal experiments. Here researchers make inferences from what happens to  $X$ s (some non-human CAM) inside the laboratory to  $Y$ s (humans) outside the laboratory. Consequently this is not straightforward causal reasoning, not even probabilistic causal reasoning.

To put it differently, biomedical experiments on animals are doubly probabilistic. Experimenters discover that some percentage of  $X$ s (the chosen animal species) in the laboratory react in some particular way and conclude that it is probable or likely that some percentage of  $Y$ s (humans) will react similarly outside the laboratory. Thus, there is probabilistic causality within the (non-human) laboratory population, probabilistic causality within the human population outside the laboratory, and an uncertainty about whether the results observed in the non-human animal population will be (statistically) relevant to the human biomedical phenomena of interest. Exactly how this latter uncertainty is characterized is crucial in determining the adequacy of animal models as CAMS.

To help characterize this uncertainty, we shall take guidance from David Hull's statement about the aim of CAMs, at least when the model and the object modelled differ (p. 105):

In reasoning by analogy, the behaviour of a poorly understood system is assimilated to the behaviour of a well-understood paradigm system. Hopefully the principles that govern the behaviour of the paradigm system can be extrapolated to the poorly known system.

To put it more formally, CAMs fit the following schema of all analogical arguments: *X (the model) is similar to Y (the subject being modelled) with respect to properties [a, ..., e]. X has additional property f. While f has not yet been observed directly in Y, it is likely that Y also has the property f.* Since CAMs are a subspecies of analogical arguments in which (some of) the premises and conclusions involve *causal analogical* claims, the CAMs must satisfy two further conditions especially relevant to its causal dimensions: (1) *the common properties [a, ..., e] must be causal properties, which* (2) *are causally connected with the property f we wish to project – specifically, f should stand as the cause(s) or effect(s) of the features [a, ..., e] in the model.*

These are rigorous requirements. But not yet rigorous enough. To determine the certainty or the probability of extrapolations from animal test-subjects to humans, we must be confident that the causal mechanisms under investigation in the non-human animal are relevantly similar to functionally analogous human mechanisms. For investigators like Bernard that assumption was innocent enough. He thought (p. 115) ‘all animals may be used for physiological investigations, because with the same properties and lesions in life and disease, the same result everywhere recurs’. Bernard did indeed recognize species differences, but he conceptualized these as being primarily quantitative – e.g., differences with respect to body weight – so that once allowance had been made for such differences in physiological formulae, the ‘same cause, same effect’ principle would hold true, and the same result would everywhere recur. This is made particularly clear in his discussion (p. 180) of differences between toads and frogs with respect to toad venom (see our 1994b). But the history of medicine makes it evident that such an assumption is not innocent.

Hence this should not be an unstated assumption. It should be an explicit condition which must be satisfied if causal inferences from non-human animals to humans are to be legitimate. That is, if animal subjects are to be good CAMs of some human biomedical phenomenon, then in addition to conditions (1) and (2) above, we must also require that (3) *there must be no causally relevant disanalogies between the model and the thing modelled.* To the extent that there are no (or insignificant) causal disanalogies between the test subjects and humans, the additional layer of probability or uncertainty mentioned earlier will be minimal. To the extent that there are important disanalogies, this additional layer

of probability will be sufficient to attenuate our confidence in animal test subjects as CAMs of human biomedical phenomena.

2. *Some reasons for thinking condition (3) is not satisfied*

Researchers who think non-human animals are good CAMs of human biomedical phenomena believe human and non-human animal systems are causally similar because they are functionally similar. Lungs oxygenate the blood, while livers remove impurities from it, whether the animal is a rat, a bird, or a human. As Lubinski and Thompson explain (p. 628; our italics):

Darwin's work suggested that to gain biological insight into human beings it may be more illuminating to study non-human animate systems rather than the inanimate models of da Vinci and Descartes. Claude Bernard, the founder of experimental medicine, used dogs in laboratory preparations as models of human physiology, assuming basic continuity in *physiological functions* across species. Both Darwin and Bernard argued that anatomy, physiology and behaviour not only look similar in different animals but often share common evolutionary origins and current regulatory mechanisms.

Bernard himself expressed similar views (p. 111):

Physiologists also follow a different idea from the anatomists. The latter, as we have seen, try to infer the source of life exclusively from anatomy; they therefore adopt an anatomical plan. Physiologists adopt another plan and follow a different conception; instead of proceeding from the organ to the function, they start from the physiological phenomenon and seek its explanation in the organism.

According to some researchers the connection between function and causal mechanisms is so tight that for purposes of biomedical research humans and non-human animals are virtually interchangeable. This idea is expressed succinctly in the standard toxicology text, *Casarett and Doull's Toxicology*:

the effects produced by the compound in laboratory animals, when properly qualified, are applicable to humans. This premise applies to all of experimental biology and medicine (Klassen and Easton p. 31; by 'proper qualification' they mean that researchers should make allowances for quantitative differences in body weight, surface area, etc.).

Other researchers might suggest that the distinction between causal mechanisms and functional properties is just a matter of description, and hence similarity of physiological function implies similarity of causal mechanism. After all, functional properties are indeed effects of under-

lying causal mechanisms, and as Sir Isaac Newton put it (p. 398): 'Therefore to the same natural effects we must, as far as possible, assign the same causes. As to respiration in a man and in a beast; the descent of stones in Europe and in America.'

True, a physiologist might describe the operations of the liver in causal terms (e.g., the mechanisms whereby it removes a foreign substance from the blood) or in functional terms (as purifying the blood), depending on the purpose in hand. Nevertheless we should not infer that two functionally similar systems have the same underlying causal mechanisms. This point is well recognized by bench scientists who are primarily interested in an organism's causal mechanisms – even when the effects of the mechanism can be described, for other purposes, in functional terms. That is, they are more interested in the liver's mechanisms for purifying blood than in the simple functional fact *that* it purifies blood. Cures and preventative strategies typically hinge on an understanding of causal mechanisms.

Pragmatists have argued that a machinist is able to repair an engine without understanding the theory behind its operation. But teratology is concerned with more than repair. One of the major objectives is to anticipate risks before they materialize. The anticipation of teratic risks in today's rapidly changing environment becomes an endless succession of screening tests unless a knowledge of mechanisms can lead to extrapolations, generalizations and shortcuts that will simplify the task. Furthermore the use of animal tests for evaluation of human risk will become more than empirical only when the degree of comparability of mechanisms between test animal and man is understood. Finally, with a better knowledge of mechanisms, unknown causes may be more easily recognized (Wilson p. 72).

That functional similarity does not imply underlying causal similarity should be apparent. Even in simple mechanical systems, like clocks, the same function can be achieved by a variety of different mechanisms. This is still more apparent in complex, dynamical systems, like biological systems. As Burggren and Bemis comment (p. 194):

The peribronchial lungs of birds, ventilated in a unidirectional fashion using a series of air sacs, and the alveolar lungs of mammals, ventilated in a tidal fashion using a diaphragm, differ considerably in structure and mechanism. Yet, both ultimately produce the same effect – full oxygen saturation of the arterial blood.

Or consider the metabolization of phenol, mentioned earlier. Phenol is metabolized by a conjugation reaction with either glucuronic acid or sulphate. The purpose of this reaction is to enhance its water solubility and thereby ease excretion. Cats, rats, pigs and humans are functionally

similar: they can all metabolize phenol. However, the mechanisms of phenol metabolism vary widely from species to species. The ratio of conjugation with sulphate to glucuronidation in humans is 80% : 12%; in rats it is 45% : 40%. By contrast, pigs are deficient with respect to conjugation with sulphate and cats are deficient with respect to glucuronidation. Wide species-variation in the mechanisms of metabolism is also seen in other compounds, like amphetamines and benzodiazepines (see Caldwell 1980 pp. 94–106). As the preceding considerations make apparent, functional similarity does not guarantee underlying causal similarity, nor does it make such similarity 'probable'. To assume it does is to commit what we term the *modeller's functional fallacy*.

Some researchers also claim that, since animals and humans are phylogenetically continuous, we can legitimately assume condition (3) is satisfied. However, even when species are phylogenetically close, as are the rat and the mouse, we cannot assume that the two species will react similarly to similar stimuli. Tests for chemically induced cancers in rats and mice yield the same results for only 70% of the substances tested (see Lave *et al.*). The figure drops to 51% for site-specific cancers (Gold *et al.* p. 245). And primates, our 'closest' biological relatives – and presumably the ideal test subjects – have biological subsystems which are significantly disanalogous with those in humans.

Non-human primates offer the closest approximation to human teratological conditions because of phylogenetic similarities.... However, a review of the literature indicates that except for a few teratogens (sex hormones, thalidomide, radiation, etc.) the results in non-human primates are not comparable to those in humans (Mitruka *et al.* pp. 467–8).

Phylogenetic continuity, even relative phylogenetic 'closeness', does not guarantee that relevant subsystems are similar in causally relevant respects. Still less does it guarantee that the interactions between those subsystems are identical. Consider the phenol case mentioned earlier. Human mechanisms for metabolizing phenol are closer to the mechanisms in rats than to the mechanisms in pigs, despite the fact that humans are phylogenetically closer to pigs than to rats. And the carcinogenic effect of aflatoxin B is more similar in rats and monkeys than in rats and mice (see Vainio *et al.* p. 20). Thus, as the preceding considerations suggest, to reason that phylogenetic continuity implies underlying causal similarity is to commit what we term the *modeller's phylogenetic fallacy*.

Admittedly, one might think that the underlying causal mechanisms are *vaguely similar*. But given the nature of biomedical phenomena and the need for detailed and exact information about causal mechanisms, vague similarity is of little help – at least for CAMs. For, as Caldwell

claims (1980 p. 106), the biomedical significance of very small differences between test-subjects produces substantially different results:

It has been obvious for some time that there is generally no evolutionary basis behind the particular drug metabolizing ability of a particular species. Indeed, among rodents and primates, zoologically closely related species exhibit markedly different patterns of metabolism.

In summary, the empirical evidence to hand gives us reason to think condition (3) is not satisfied. These empirical findings are to be expected given the theory of evolution.

### *3. Evolutionary explanations for the failure to satisfy condition (3)*

We argued earlier (1993b) that the theory of evolution gives us reason to think that animal models are good HAMs of human biomedical phenomena. Our common evolutionary heritage suggests we all have some mechanisms for oxygenating the blood, for regulating body temperature, for reproducing, etc. – that is why animal models are good HAMs. But the issue now is not whether animals are good HAMs, but whether they are good CAMs. For some narrow purposes, perhaps animal models are good CAMs. For instance, were we merely examining the gross effects of concentrated sulphuric acid on animal tissue, we could probably ignore species differences. These effects depend on fairly low-level chemical properties studied by organic chemists.

However, most biomedical phenomena of interest depend more on the properties between evolved systems and subsystems. If we are interested in, say, the long-term effects of exposure to low levels of sulphuric acid (perhaps from acid rain) we shall doubtless find different species react differently to the very same stimuli. Indeed, we shall encounter intra-specific variation too.

Certainly that is what the theory of evolution would suggest. For functional similarities are often not supported by the same causal mechanisms. ‘Descent with modification’ means, in part, ‘modification of anatomical and physiological subsystems, and the relations between them’. Evolution creates biological systems which are hierarchically complex.

We animals are the most complicated things in the known universe.... Complicated things, everywhere, deserve a very special kind of explanation. We want to know how they came into existence and why they are so complicated. The explanation, as I shall argue, is likely to be broadly the same for complicated things everywhere in the universe; the same for us, for chimpanzees, worms, oak-trees and monsters from outer space. On the other hand, it will not be the same for what I shall call ‘simple’ things, such as rocks, clouds,

rivers, galaxies and quarks. These are the stuff of physics. Chimps and dogs and bats and cockroaches and people and worms and dandelions and bacteria and galactic aliens are the stuff of biology (Dawkins p. 1).

Biological objects differ from rocks and stars because of their *structural organizational complexity*. Humans are not 'essentially' different from rats, nor are we 'higher' life-forms. But we are differently complex. DNA itself exhibits such complexity, and produces further complexity in anatomy and physiology. Genes do not do their work one by one. Rather they 'conspire' to produce effects at the cellular level – and ultimately affect the whole organism. Thus, physiological effects are produced not only by the genes, but by evolved complex relations between them.

To put it differently, most biomedically significant properties are relational properties – properties dependent on the interaction of the organism's subsystems. Many of these properties are emergent properties arising from evolved hierarchical organization: biological entities at lower levels of organization are compounded to produce biological entities at higher levels of organization (macromolecules to cells, cells to tissues, tissues to organs, organs to organisms). As Ernst Mayr comments (p. 15), in rejecting the mechanistic atomism of the old physiology texts:

Systems at each hierarchical level have two properties. They act as wholes (as though they were a homogeneous entity), and their characteristics cannot be deduced (even in theory) from the most complete knowledge of the components, taken separately or in other combinations. In other words, when such a system is assembled from its components, new characteristics of the whole emerge that could not have been predicted from a knowledge of the constituents.... Indeed, in hierarchically organized biological systems one may even encounter downward causation.

Resultant species differences are biologically significant. 'The species is one of the basic foundations of almost all biological disciplines. Each species has different biological characteristics' (Mayr p. 331). Species differences, even when small, often result in radically divergent responses to qualitatively identical stimuli (e.g., the details of phenol metabolism discussed earlier). Evolved differences in biological systems between mice and men cascade into marked differences in biomedically important properties between the species.

That is, although two species may share a common stock of biochemical parts, the organizational structure of those parts can lead to different biological reactions. The effects of these differences undermine any confidence that condition (3) is satisfied. Satisfaction of condition (3) cannot be assumed *a priori* – it can only be established empirically. That is, we can be confident that condition (3) is satisfied only after we have

conducted tests on humans – and yet animal tests are deemed desirable primarily to eliminate the need for such tests on humans.

In summary, there are compelling theoretical and empirical reasons for suspecting that condition (3) is not satisfied. That is, there is evidence that the same stimuli produce different responses, either because significant subsystems are different, or because their mutual interactions are different, or both. In other words, the analogies between animal CAMs and the human systems they model are frequently weak.

#### 4. *Weak CAMs*

Many researchers recognize that animal test subjects are causally dissimilar to the human systems they are intended to model.

It is the actual results of teratogenicity testing in primates which have been most disappointing in consideration of these animals' possible use as a predictive model. While some nine subhuman primates (all but the bushbaby) have demonstrated the characteristic limb defects observed in humans when administered thalidomide, the results with eighty-three other agents with which primates have been tested are less than perfect. Of the fifteen listed putative human teratogens tested in non-human primates, only eight were also teratogenic in one or more of the various species.... The data with respect to the 'suspect' or 'likely' teratogens in humans under certain circumstances were equally divergent. Three of the eight suspect teratogens were also not suspect in monkeys or did not induce some developmental toxicity (Schardein pp. 20–3).

However, most of these researchers claim that animals serve as good CAMs of human biomedical phenomena, despite their (possibly numerically small) causally significant disanalogies. They believe that humans and animal test subjects share enough biomedically significant causal mechanisms to justify inferences from animals to humans. Or, to use the language of the formal analysis offered earlier, experimenters assume animal models only violate condition (3) partially.

What exactly does this mean? This is difficult to know. But here is one interpretation. Begin with two systems,  $S_1$  and  $S_2$ .  $S_1$  has causal mechanisms [a,b,c,d,e],  $S_2$  has mechanisms [a,b,c,x,y]. When stimulus  $s_f$  is applied to subsystems [a,b,c] of  $S_1$ , response  $r_f$  regularly occurs. We can therefore infer that were  $s_f$  applied to subsystems [a,b,c] of  $S_2$ , it is highly probable that  $r_f$  would occur.

This would be a plausible inference, however, only if the common mechanisms [a,b,c] are independent of the differing mechanisms [d,e] and [x,y]. If any of the common mechanisms are dependent on the different mechanisms, then their interactions may well produce diver-

gent responses to qualitatively identical stimuli. The extent of the divergence will depend not merely on the numerical extent of the different mechanisms, but primarily on the strength of the relation of dependence and the qualitative significance of the different mechanisms. Given the empirical and theoretical evidence adduced in the previous sections, we have good reason to expect that there will be significant disanalogies which will undermine the strength of the inference from  $S_1$  to  $S_2$ .

Many researchers have tried to avoid the problem of species disanalogies by appealing to scaling formulae (a method hinted at in the *Casarett and Doull* passage quoted on p. 148 above). These researchers do not deny that animal models are weak analogues:

While reasonably exact extrapolations can be made between different physical objects such as from a small circle to a large circle, the assumption of strict biological similitude between animals, especially those of different species, is unrealistic and only approximations are achieved (Calabrese p. 501).

However, they do claim we can accommodate species differences purely quantitatively. The theoretical basis for this accommodation is explained by Calabrese (*ibid.*):

Despite this fundamental limitation, it is recognized that in the animal kingdom surface area per unit weight decreases with increasing body weight and basal metabolism per unit weight declines with increasing body weight. Such relationships provide the basis for the organism developing a constant ratio of heat production to the external surface area.

That is, toxicologists recognize that species differences can arise from qualitative differences, either from (1) differences in pharmacokinetics (affecting doses delivered to target tissues) or (2) inter-specific differences with respect to tissue sensitivity (see Klassen and Easton). It is hoped to accommodate these differences by making quantitative adjustments based on body weight and external surface areas. In this way qualitative differences are treated as quantitative differences which can be compensated for in physiological formulae. As Calabrese comments (p. 501):

It suggests that numerous structural and physiological parameters are mathematical functions of body weight. This appears to be true regardless of species.

For some biological characteristics this is a reasonable assumption. Doubtless the strength of an animal's supporting structures or the weights of its organs can be estimated in these ways. Moreover, the rates at which some physiological functions are achieved may well be related to body weight or surface area. However, metabolic differences between

species, which are centrally important in toxicological and teratological investigations, are not merely relative to size.

Even where distinct species achieve similar functions at distinct rates related to differing body size, we cannot infer that the mechanisms underlying these functions are similar. For example, we know that there are at least seven metabolic pathways unique to primates (Caldwell 1992). And even where the pathways are similar, there is considerable inter-specific variation with respect to the extent of various reactions. As one widely used pharmacology text sums it up: 'The lack of correlation between toxicity data in animals and adverse effects in humans is well known' (Goth p. 46).

This lack of correlation stems from the extreme sensitivity of biological phenomena to even very small species differences (see our 1994a). Biomedical phenomena may vary radically even among different strains of the same species. Such differences are evident in the reaction of non-human animals to thalidomide:

An unexpected finding was that the mouse and rat were resistant, the rabbit and hamster variably responsive, and certain strains of primates were sensitive to thalidomide developmental toxicity. Different strains of the same species of animals were also found to have highly variable sensitivity to thalidomide. Factors such as differences in absorption, distribution, biotransformation and placental transfer have been ruled out as causes of the variability in species and strain sensitivity (Manson and Wise p. 228).

##### 5. *Instrumentalism: 'It just works'*

As a last line of defence researchers might contend that we know animals are causally similar to humans by experience. As Giere claims (p. 233), 'As for the statement that humans are not rats, that is obviously true. But of the approximately thirty agents known definitely to cause cancer in humans, all of them cause cancer in laboratory rats – in high doses.'

This 'fact' appears to lend considerable credence to the claim that animals are reasonably good models of human biomedical phenomena. As it turns out, however, Giere's claim is seriously misleading. None the less his assertion provides an ideal opportunity to state the 'it just works' argument more precisely, and to scrutinize more carefully the status of carcinogenicity testing.

According to the International Agency for Research on Cancer there are 26 (of 60,000) chemicals which have been shown to be carcinogenic in humans. Giere's claim is misleading because the *usefulness* of a test is

a function not just of the *sensitivity* of the test (the proportion of human carcinogens that are carcinogenic in rats), but also of the *specificity* of the test (the proportion of human non-carcinogens that are non-carcinogenic in rats). Research indicates that the specificity of such tests is quite low. In one test, rats developed cancers when exposed to 19 out of 20 probable human non-carcinogens. If so, specificity may be as low as 0.05 (Lave *et al.* 1988 p. 631).

The drawbacks of animal carcinogenicity tests can also be seen in a slightly different way. It is known that the relationship of *concordance* between rats and mice (the agreement in test outcomes for non-site-specific tumours – both positive and both negative) is 70%. Current testing policy assumes that the relation of concordance between rats and humans is the same as that between rats and mice – a highly questionable assumption indeed. In fact, were this a reasonable assumption, it would undercut the researchers' contention that phylogenetic continuity implies similarity of causal mechanisms.

However, even if we granted this questionable assumption, using rodent studies to assess human cancer risk would result in 30% of the tested chemicals being misclassified (3% false negatives and 27% false positives). The social cost of such misclassification is enormous (Lave *et al.* 1988 pp. 631–2). Evidence demonstrating the limitations of animal tests for carcinogenicity has become so overwhelming that even governmental agencies are beginning to adopt new non-animal-based research strategies (see Brinkley; also Vainio *et al.* pp. 27–39). Despite this factual error, Giere has posed an interesting argument which merits consideration. Apparently the argument would go something like this. If (a) species S has produced biomedical reactions (developed cancer, etc.) upon receiving some type of stimulation (e.g., chemical stimulation)  $x\%$  of the time, and if (b) humans respond in a similar way to identical chemical stimulation  $y\%$  of the time, and if (c) there is a strong correlation between the response rates in the experimental population (members of species S) and the response rates in the exposed human population – suppose their responses are concordant  $z\%$  of the time (e.g., 80% of the time), then if that species reacts to some new but related stimulus, we can infer that there is an (approximately)  $z\%$  probability that the human will react similarly.

Is this a plausible expectation? We do not see how. Suppose scientists knew rats and humans respond in the same ways to potential carcinogens 80% of the time. Even so, the inference that we would find similar responses to other potential carcinogens would be viable only if we had reason to think that the sample class was representative of all carcinogens. For example, if twenty-four of the thirty carcinogens mentioned

by Giere were members of a sequential series of hydrocarbons (hydrocarbons with a common general formula – e.g., the paraffin series whose members are instances of the general formula  $C_nH_{2n+2}$ ), then what we may have discovered is that rats and humans react similarly to a certain subclass of chemicals. Furthermore, since we have not identified the carcinogenic potential of many chemicals, nor many of the mechanisms which often produce cancer, how could we be reasonably confident that the sample class was representative?

More generally, most animal research is not aimed at making crude predictions, but at uncovering causal mechanisms of particular human conditions (heart disease, the course of cancer, Parkinson's disease, etc.). However, as we noted earlier, a CAM is serviceable only if the mechanisms of the animal's disease or condition are, in fact, *causally* similar to those in humans. We can have reason to believe they are causally similar only to the extent that we have detailed knowledge of the condition in *both* humans and animals. However, once we have enough information to be confident that the non-human animals are causally similar (and thus, that inferences from one to the other are probable), we likely know most of what the CAM is supposed to reveal. So the value of a CAM, even in this idealized case, is less than we might expect.

One final attempt to defend the 'it just works' argument goes as follows: surely we just know, from surveys of primary research literature and from histories of medicine, that the institution of animal research is a powerful source of biomedically significant information about humans (see Smith and Boyd pp. 25–9). Notice, though, that this response is offered not only as a defence of using animals as CAMs, but also of using them as HAMs. Hence even if the primary literature did reveal that animal experiments were an important source of information about humans, it would be difficult to extract the historical role that animal CAMs played from the role that animal HAMs played. That is, even if this claim were true, it would not provide any special reason for thinking that animal models were good CAMs, or, in other words, for thinking that condition (3) is satisfied.

Setting that worry aside, this means of determining the success of animal experimentation is not as straightforward as it might seem. For it is likely that both sources of information will report failures of research, but will *seriously under-report* manifest dissimilarities between animals and humans. If a researcher is trying to discover the nature of human hypertension, and conducts a series of experiments on an animal only to discover that the animal cannot develop hypertension, those findings are not likely to be reported. And, even if negative findings are

sometimes reported, they are less likely to be read and discussed by professionals – especially if the negative result does not uncover significant data to account for that failure. For this reason, it is misleading to assess the fecundity of the institution of animal experimentation simply by tallying successes, or even ratios of successes to failures, in the extant research literature.

There are further difficulties in documenting the success of animal experimentation by reading standard ‘histories’ of great events in biomedical research. When historians of medicine discuss the history of some biomedical advance, it is not unusual for failed experiments to be under-reported (even when they appear in the primary research literature). Historians report only events crucial to understanding the current state of the science. Failed experiments (usually vital to the actual development of science) are often under-reported (perhaps, even, because of their ubiquity) in these histories.

This is not to question either the factual accuracy of reports by or the integrity of medical historians. Anything but. Careful studies of the history of medicine can be quite instructive. The question here is not the accuracy of the reported facts, but how those facts are interpreted. That is, given the human tendency to rewrite even our personal histories in light of our present beliefs (Ross pp. 342–4) it would be surprising if medical historians did not write the history in a way that articulates their understanding of that science. Since the use of non-human animals as CAMs is integral to the current view of the biomedical sciences, we should not be surprised to find that these histories often emphasize the ‘successes’ of non-human CAMs. This gives us one further reason to think that it is no simple matter to come up with the sort of evidence required to substantiate the ‘it just works’ argument. More is required than counting ‘successes’ reported in the literature.

## CONCLUSION

Researchers claim that non-human animals can be used as CAMs to uncover underlying causal mechanisms of human disease. We disagree. We have argued that animal tests are unreliable as tests to determine the causes and properties of human disease. Available evidence and the theory of evolution lead us to expect that evolved creatures will have different causal mechanisms undergirding similar functional roles.

On the other hand, there are good theoretical reasons to think animals can serve as HAMs. The theory of evolution leads us to expect that phylogenetically close species will have numerous functional similarities. HAMs work primarily because of similarities in functional properties.

Thus, animal research may be more viable in the context of basic research, when we have relatively little knowledge of biological mechanisms, than in goal-orientated research which seeks to use animals as causal 'test beds' for hypotheses about human biomedical phenomena.

This conclusion fits well with our reading of the history of biomedicine. Those cases standardly offered as demonstrating the benefits of animal experimentation (poliomyelitis, insulin, etc.) used animal tests as HAMs – not CAMs. That is, they are cases where uses of animals in basic research prompted insights which ultimately lead to new understanding of or treatments for human disease.

Of course, if animal models can serve as HAMs to spur research, perhaps clinical investigations, cell cultures, computer simulations, or epidemiological studies might be effective HAMs as well. At the very least, these other methods need no longer be construed as poor cousins to animal research. They may all become a more important part of basic biomedical research.<sup>2</sup>

*East Tennessee State University*

## REFERENCES

- AMA White Paper 1988: *The Use of Animals in Biomedical Research: the Challenge and Response* (American Medical Association).
- Bernard, C. 1949: *An Introduction to the Study of Experimental Medicine* (Paris: Henry Schuman).
- Brinkley, J. 1993: 'Animal Tests as Risk Clues: the Best Data May Fall Short', *New York Times National*, 23 March 1993, C1, C20–1.
- Burggren, W.W. and Bemis, W.E. 1990: 'Studying Physiological Evolution: Paradigms and Pitfalls', in M.H. Nitecki (ed.), *Evolutionary Innovations* (Univ. of Chicago Press).
- Calabrese, E.J. 1983: *Principles of Animal Extrapolation* (New York: Wiley).
- Caldwell, J. 1980: 'Comparative Aspects of Detoxification in Mammals', in W. Jakoby (ed.), *Enzymatic Basis of Detoxification*, Vol. 1 (New York: Academic Press).
- 1992: 'Species Differences in Metabolism and their Toxicological Significance', *Toxicology Letters* 64–5, pp. 651–9.
- Dawkins, R. 1987: *The Blind Watchmaker* (New York: Norton).
- Giere, R. 1991: *Understanding Scientific Reasoning* (New York: Harcourt Brace Jovanovich).
- Gold, L. et al. 1991: 'Target Organs in Chronic Bioassays of 533 Chemical Carcinogens', *Environmental Health Perspectives* 93, pp. 233–46.
- Goth, A. 1981: *Medical Pharmacology: Principles and Concepts*, 10th edn (St Louis: C.V. Mosby).
- Hempel, C.G. 1965: *Aspects of Scientific Explanation* (New York: Macmillan).
- Hull, D. 1974: *Philosophy of Biological Science* (Engelwood Cliffs, N.J.: Prentice-Hall).
- Klassen, C. and Easton, D.M. 1993: 'Principles of Toxicology', in *Casarett and Doull's Toxicology*, 4th edn (New York: McGraw Hill), pp. 12–49.
- LaFollette, H. and Shanks, N. 1993a: 'Animal Models in Biomedical Research: Some Epistemological Worries', *Public Affairs Quarterly* 7, pp. 113–30.

<sup>2</sup> We thank philosophers George Gale, James Rachels, Mark Parascandola and Dale Cooke, biologists Andrew J. Petto and Rebecca Pyles and medical historian John Parascandola for helpful comments on and criticisms of drafts of this paper.

- 1993b: 'The Intact Systems Argument: Problems with the Standard Defense of Animal Experimentation', *Southern Journal of Philosophy* 31, pp. 323–33.
- 1994a: 'Chaos Theory: Analogical Reasoning in Biomedical Research', *Idealistic Studies* 24, pp. 241–54.
- 1994b: 'Animal Experimentation: the Legacy of Claude Bernard', *International Studies in the Philosophy of Science* 8, pp. 195–210.
- Lave, L.B. *et al.* 1988: 'Information Value of the Rodent Bioassay', *Nature* 336, pp. 631–3.
- Lubinski, D. and Thompson, T. 1993: 'Species and Individual Differences in Communication Based on Private States', *Behavioral and Brain Sciences* 16, pp. 627–80.
- Manson, J. and Wise, L.D. 1993: 'Teratogens', in *Casarett and Doull's Toxicology*, 4th edn (New York: McGraw Hill), pp. 226–81.
- Mayr, E. 1988: *Toward a New Philosophy of Biology* (Harvard UP).
- Mitruka, B.M. *et al.* 1976: *Animals for Medical Research: Models for the Study of Human Disease* (New York: Wiley).
- Newton, I. 1687: *Principia*, Vol. 2, trans. Motte (Berkeley: Univ. of California Press, 1962).
- Ross, M. 1989: 'Relation to Implicit Theories in the Construction of Personal Histories', *Psychological Review* 96, pp. 341–57.
- Salmon, W. 1984: *Scientific Explanation and the Causal Structure of the World* (Princeton UP).
- Schardein, J.L. 1985: *Chemically Induced Birth Defects* (New York: Marcel Dekker).
- 'Sigma Xi Statement on the Use of Animals in Research' 1992: *American Scientist* 80, pp. 73–6.
- Silverstein, A. 1989: *A History of Immunology* (New York: Academic Press).
- Smith, J. and Boyd, K. 1991: *Lives in the Balance: the Ethics of Using Animals in Biomedical Research* (Oxford UP).
- Vainio, H. *et al.* 1992: *Mechanisms of Carcinogenesis in Risk Assessment* (Lyon: International Agency for Research on Cancer).
- Wilson, J.G. 1977: 'Current Status of Teratology', in J.G. Wilson and F.C. Fraser (eds), *Handbook of Teratology*, Vol. 1 (New York: Plenum Press).