The Humane Society Institute for Science and Policy
Animal Studies Repository

6-2004

Animal Model Research: The Apples and Oranges Quandary

Kenneth J. Shapiro
Psychologists for the Ethical Treatment of Animals, ken.shapiro@animalsandsociety.org

Follow this and additional works at: http://animalstudiesrepository.org/acwp_arte

Part of the Animal Experimentation and Research Commons, Animal Studies Commons, and the Design of Experiments and Sample Surveys Commons

Recommended Citation
Shapiro, K. J. (2004). Animal model research: the apples and oranges quandary. ATLA-NOTTINGHAM-, 32, 405-410.

This Article is brought to you for free and open access by the Humane Society Institute for Science and Policy. It has been accepted for inclusion by an authorized administrator of the Animal Studies Repository. For more information, please contact eyahner@humanesociety.org.

Animal Model Research: The Apples and Oranges Quandary

Kenneth J. Shapiro

Psychologists for the Ethical Treatment of Animals, 403 McCauley Street, Washington Grove, MD 20880, USA
E-mail: *kshapiro@psyeta.org*

Summary — In this paper, I explore the premises underlying the problem of the evaluation of animal models. I argue that the presence of similarities and differences between the model and the modelled, although historically and currently a dominant antinomy framing evaluation, is not a bottom-line consideration. What is critical is 1) whether we learn and 2) whether we improve treatment through the animal model research. Similarity between model and modelled and the closely related concept of validity are not coterminus with these critical evaluative measures. In fact, differences between the model and modelled also can provide impetus to new understanding and treatment innovations. The apples and oranges argument — that model and modelled are incommensurable — whether based on theoretical or empirical grounds, is not an adequate critique. Continuing argument that relies heavily on similarities versus differences is unconstructive, reducing to the proverbial half-empty/half-full bottle argument.

Key words: *animal models, evaluation, validation.*

Introduction

In my view, science-based arguments are less well developed than are ethics-based arguments in the debate over animal research. At the 3rd World Congress, I presented a paper on the problem of the scientific evaluation of animal models (1). In it, I critically reviewed the history of evaluative methods, outlined a method using standard social scientific methods, and gave some examples of its application (2). My effort here is to deepen the science arguments through a critique of the present strategy at the level of the concepts that presently frame it.

A great deal of the debate over animal research centres on whether the model and modelled are best thought of as similar or different: can they be sufficiently similar to allow extrapolation from model to modelled; to the degree that they are different, does not that invalidate a model; if they are similar, does not that fact preclude our use of them?

The present thesis is that a similarity/difference (S/D) calculus is not the bottom-line consideration in the evaluation of animal models of biomedical and psychological research. What is critical is whether we learn (the basic science question) and whether we improve treatment through the use of models (the applied science question). These are not necessarily directly related to validity and similarity or inversely related to differences and invalidity. More critically, animal models can and must be assessed independently of S/D.

On a personal note, my exploration of S/D as an organising frame began a decade ago, when writing a book on the evaluation of animal models (3). Although I approached that work with the view that S/D was a prospective cleaver, I found that it does not work either on a conceptual or empirical level. In fact, a review of the history of the animal research debate revealed that the S/D frame contributed to the perpetuation of the largely polarised state of the controversy. This led me to adopt standard social scientific evaluative methods, such as citation analysis, survey of relevant users, measures of pain and treatment evaluation studies. I believe the use of these methods puts the science argument on a stronger footing, both in the evaluation of any particular proposal and in the larger debate over the appropriateness and effectiveness of the animal model strategy generally.

A disclaimer: I am talking about basic and applied biomedical and psychological research here, not toxicity testing. As the purpose and procedures in the two enterprises are different, so is the meaning and locus of S/D and validation, as I will clarify.

I begin with some introductory material to 1) establish the concepts and historical backdrops of S/D and 2) show that indeed S/D is a primary frame within which people think about animal research.

Discussion

People use the titular cliché, "like comparing apples and oranges", to argue that two items are so different as to be incomparable. If the apple were a model of the orange, could we learn anything about oranges from the study of apples, the basic science

question; and, the applied science question, could we learn to grow them better.

Upon reflection, of course, apples and oranges are alike in many ways. On a gross morphological level, both are round and of similar size, have skins, seeds and pulp; both grow on trees and have similar stages in development. On a more abstract and conceptual level, both are fruit, edible, and good for you. On a molecular level, both are chemically acidic, largely consist of water, and have similar cellular structure. They are a lot alike and a lot different.

More generally, for any two things that are not identical, there exists a set of similarities and differences. That is why metaphoric or analogical talk is so rich and evocative — a person is a banana. Pairings vary in the degree of S/D. We can construct a meaningful metaphor from virtually any pairing. Many, but not all, such pairings give us insights into the meaning of the target object for which an analogy has been constructed. The mind is a computer.

Philosophers use the term "incommensurable" to refer to two things so different that any comparison is misleading. Is a typical animal model and target incommensurable? Unfortunately, for those opposing animal research, they are not. For a typical animal model/modelled, similarities obtain relevant to the purpose of drawing the comparison. But, fortunately, and more importantly, I will show that such similarities are not critical in the evaluation of the model. In any case, how can we judge this? How much are apples and oranges similar and how much different? Are they more similar than different? Are the similarities more critical in any gains made from the model?

I am arguing that measures of increased understanding and more effective treatment are critical to the evaluation of animal models, while measures of the degree of similarity to differences are not, and that these two sets of measures are not related in any simple way.

History of S/D in the comparison of humans and other animals

A major theme in Western thought is the concept of human being: what is a "man", or a "person"? Animals other than humans provide an important categorical foil in the evolution of this concept. A brief survey of changes in the concept of human being reveals shifts from a view that emphasises differences between humans and other animals (H/A), to one that emphasises similarities, to a mixed or ambivalent view, to, again, similarities as focal.

In the Western classical tradition, both Greco-Roman and Judeo-Christian thinkers described a discrete categorical difference between H/A — the first based on the possession of rationality, the second of a soul. Descartes continued this traditional categorical cleavage, opposing the human rationality/soul to the machine-like nature and thus radical differentness or otherness of animals other than humans.

In his theory of evolution, Darwin shifts from an emphasis on discontinuity to continuity. Beneath the apparent differences in the panoply of life forms, he finds a common ancestry. Although Darwin also uses differences to define species, his scientific style of thought unifies apparently disparate phenomena under general principles, to emphasise similarity and continuity. Humans are made in the image of animals other than humans, not God. The contemporary synthesis of Darwinian Theory and genetics also emphasises continuity by demonstrating similarity at the biochemical level, in addition to the anatomical, physiological, and behavioural levels.

This mixed or ambiguous modern scientific view is reinforced by the complexity of contemporary constructions of animals, such as the "pet", in which animals connote both human-like members of the family and vestiges of our bestial ancestors.

Arguably, we are on the cusp of a movement back to an emphasis on similarity through the influence of biocentrism, the ecological viewpoint, theories of systemic relations (chaos theory), cognitive ethology and genetic engineering. Together, these developments discover and literarily create animals that are more like us in intellective and social capabilities and in physiological function.

Singer's scientists' dilemma

Singer exploits the perennial antinomy of S/D in his analysis of the dilemma facing researchers (4). He argues that scientists emphasise similarity to justify their research, but, implicitly, difference to justify the use of non-human animals in procedures to which they would not subject humans.

Unfortunately, the incoherence cuts the other way in the arguments of animal advocates. Anti-vivisectionists argue difference to disqualify the experiment (the science argument), but use similarity to argue equal consideration (the ethics argument). Interestingly, its precursor, the 19th century humane movement posited difference as the basis of obligation of kindness to non-human animals — only humans (of a certain social class) are capable of being kind to other beings (hence the term "humane").

The concept of model in terms of S/D

As by definition any model/modelled relation is one of analogy not identity, the relation consists of a set of similarities, as well as differences. Despite this, scientists, at different times historically and in dif-

ferent contexts currently (media and political settings), have claimed that models are similar "in every respect" (5). In response, opponents of animal research refute this claim by describing respects in which a particular model is different from the condition modelled. The result is a significant waste of time and energy and a failure of constructive dialogue.

Recently, LaFollette & Shanks (6) offered the following criterion for the evaluation of an animal model: For a model to be a good model, "there must be no causal relevant disanalogies between the model and the thing modelled". On the basis of evolutionary, systems and chaos theory, they argue that, in principle, model and target cannot be identical or "isomorphic".

This is a straw person argument. One does not need these sophisticated theories to argue that the model and target are not causally identical, as it is true by definition, logic and common sense. Even a non-human animal infected with the same tuberculosis germ as a human has a different disease than the human so infected. Even an animal genetically engineered to have a compromised human immune system has a different immune system in a different physiological setting and a different disease than a human.

LaFollette & Shanks justify their extended exercise by pointing to the writings of Claude Bernard, in which he argues that identity is possible by way of a now outmoded view of the necessary relation between function and physiology. They also refer to the American Medical Association paper referenced immediately above that denies the necessity of model/modelled differences (5).

Validity and similarity

Validity is a major concept in the current debate over the effectiveness of animal models. The validity of a model refers to the degree of model/modelled similarity or its fidelity. High fidelity models claim to duplicate very closely causal mechanisms in the target disorder; low fidelity models are only loosely analogous to it. Compare a pump as a model of the workings of a heart; a telephone switchboard and, more recently, a computer as a brain; mathematical models of the nervous system; and structural models of biochemical reactions.

In toxicity testing, investigators search for a model that can be used as a substitute or stand-in for the target. They require the model to react in the same way as the target to the presence of a drug or other product in terms of potency, efficacy and toxicity. Here, high fidelity is critical. A model, to be useful, must function as a putative locus of validation — notwithstanding the fact that the drugs and products eventually are tested in the human context. By contrast, in basic and applied research, the function of the model is different and broader. The primary function of these low fidelity models is to generate hypotheses that can be tested in the target.

Validity is distinguished from reliability: the latter refers to the problem of consistency, repeatability and accuracy of measurement; the former to the degree to which what is measured is what one intended to be measured. In the model research context, reliability does not relate to similarity or fidelity of model to target, but to confidence in measurement of relevant variables in the model.

The relation between validity and the problem of evaluation of animal models is complex and often misunderstood. I present the following effort at clarification as three propositions.

1. *An animal model is never finally or fully validated.*

Validity is relative as a model and target can never be completely the same. Validity is with respect to any particular aspect or relationship of the disorder being modelled. The degree of validity of a model is a function of the accumulated aspects compared and found similar to date. Validity is ongoing and relative, and it is not a bottom-line criterion for evaluation. We never can say we have a model of depression, bulimia or cystic fibrosis in the strong sense that it can function as the locus of validation of some hypothesis about one of these disorders. In the final analysis, any hypothesis must be confirmed in the human setting. As a model is not identical with the human disorder, it is never fully or finally validated to then be a stand-in for the human disorder. The proper and potentially effective function of a model function is as a locus of discovery, not of confirmation or justification.

This is a common misunderstanding, certainly in the public's view, and one often exploited by scientists, including some who know better. We cannot and should not test an animal model to confirm a hypothesis about a human disorder.

2. *Even an animal model for which some degree of validation has been demonstrated is not necessarily a contribution to understanding or treatment effectiveness.*

Validity often is limited to the confirmation of aspects of the model that are already known to obtain in the modelled. They do not constitute new findings. Validity is only a bottom line gain when it involves an aspect validated in the target for the first time. In principle, scientists can develop a high fidelity model that teaches us nothing new about the disorder or its treatment.

3. *Most animal models in biomedical and psychological research are not validated, even in the limited sense described. I return to this below.*

The role of differences

Validity is limited as a criterion in the evaluation of animal models, as it is restricted to the identification of similarities between model/modelled. However, we also learn from differences.

An important theory of meaning in contemporary linguistics and philosophy is that meaning means through contrast. Much of what we know of "up" is through a consideration of "down", of "right" by way of "wrong", even of "vertebrate" by comparison to "invertebrate". More generally, we understand something in context — the same word or even proposition means differently in different contexts. Comparison, contrast and differentness are the stuff of understanding.

Comparative history, religion, linguistics and anatomy all provide a sophistication and depth of understanding that study of a particular history or anatomy does not. In what we seek to understand, a range of variables is at play. The comparison of differences helps expose that range. Consider individuals in one species contracting a disease that those in a second species do not. Differences between the two species are likely to be informative to an understanding and treatment of the disease. The study of differences as well as similarities generates hypotheses. Further, even a similarity disconfirmed, and thus invalidated, can be informative or suggest treatment possibilities.

A strength of the strategy of using a model — any model, whether or not animal — is that it is generative of an extensive, really inexhaustible, range of hypotheses about the modelled. On the other hand, that models have this limited generative function and that there is no necessary relation between degree of fidelity and the productiveness of that generativity has strong ethical implications that spike the move to the use of non-animal models.

The relation between model and target in practice

I am arguing that the proper use of animal models is the limited one of generating hypotheses that must be validated in a human context and that, in the evaluation of a given animal model, validity is not a critical consideration. Let me reiterate this thesis in a different way by taking a closer look at the process of the animal research enterprise. We need to distinguish among processes: a) in actual practice; b) as presented in the media and white papers by animal research advocates; and c) as they are supposed to work.

The practice in practice

My study of animal models of eating disorders demonstrated that, rather than arising from extensive familiarity with the eating disorders in its human settings, the models build on procedures already in the scientific research repertoire. More critically, once an animal model is defined, its development proceeds largely through further research consisting of recursive and, often, duplicative investigation of already studied variables and of other models within the experimental literature. In place of continuous recourse to the target phenomenon for validation of the models, it relies on confirmation by the insular criterion of consistency with other comparably insulated research. In practice, animal models are not validated, even in the limited sense we have described.

The practice as presented

Often the media and animal research advocates present the model as if it is the locus of justification — the site of validation rather than of generation of hypotheses. Their claim is that for all intents and purposes, the model is truly a stand-in for the target. Any causal process or effective treatment that obtains there, will obtain in the target. A model once developed is fully validated and can be used to test further hypotheses. In effect, the claim is that the model is an identical to the modelled.

As was the case when animal models of infectious diseases were developed by infecting the model with the same germ that affects a human, with genetic engineering, some scientists will present a model that includes inserted human genes as identical to the modelled. Although genetic engineering may increase the likelihood that the hypothesis generated will be predictive in the modelled, the subsequent causal process, disease process and effective treatment that obtains cannot be identical to that in humans. Such models must be compared to hypotheses generated by alternative models, such as human tissue or organ cultures, or direct observation of processes with high-tech imaging.

The practice as it is supposed to work

Because it cannot be identical to the target, the animal model is limited to a hypothesis generator. It is a locus of discovery not a locus of justification. Discovery is the development of hypotheses, ideas, speculations and justification is the testing of those hypotheses and ideas. It is one of many possible loci. Science traditionally has used other kinds of models to generate hypotheses, beside animals — for example, machines (most recently the computer as a model of the nervous system) and mathematical models (to

predict the molecular structure of therapeutic drugs and toxic chemicals). Of course, another locus of discovery is the target itself. In clinical practice involving the target population, discoveries are made that are not yet formally tested. They can be tested first in an animal model. This is done because scientists value lab-based animal research more than they do clinic-based research. They see the former as a more controlled setting. They give up the presence of the target phenomenon in a setting where there is less control for a setting in which they have greater control but are working only with an analogue, a model of the phenomenon. But, again, they eventually must be tested in the clinical setting, accepting whatever lesser control obtains there.

To clarify a possible point of confusion about the model as locus of discovery: the hypothesis to be tested in the target is itself a product of a scientific process of hypothesis testing. Only those variables that work in the model, at a statistical level, go on to be tested in the target.

What are the odds that it will work — that the results also will obtain in the target? What do they need to be for this strategy of animal model to be effective? Opponents of animal research, arguing the science issue, often critically claim that the prediction rate is only 50%, a rate no better than tossing a coin. This is a misunderstanding of the situation. Let's say that the prediction rate is only 50% — this means that, half the time, the hypothesis developed in the model and tested in the target will not be confirmed in the target. The causal relation of treatment effect does not reach a statistical level of 0.05 even though it did in the model. This rate is not a strong critique of the animal model process unless there is a better method of generating hypotheses. The 50% success rate must be compared, not to a coin toss, but to other ways of generating hypotheses — against both non-animal models and hypotheses from clinical observation. If these other ways only produce a 25% success rate, then clearly, animal models are a more efficient way of generating hypotheses.

These comparisons, between other non-animal models and the clinic as generators of hypotheses that are confirmed in the modelled, are the critical criteria in the evaluation of animal models. As we have demonstrated, these hypotheses issue from differences, as well as similarities.

Elsewhere I have described a scientific method by which to evaluate these critical comparisons through the use of citation analysis, survey of relevant practitioners, review of treatment evaluation literature and measures of invasiveness. I also have suggested that historical analysis of the provenance of findings supplement these social and natural scientific analytic tests. Applying and extending these methods, Dagg suggests the use of a ratio of numbers of animals to numbers of citations as a measure of effectiveness in the evaluation of animal models of cancer research, 1990–1999 (7).

References

1. Shapiro, K.J. (2000). Evaluation of animal model research. In *Progress in the Reduction, Refinement, and Replacement of Animal Experimentation* (ed. M. Balls, A-M. van Zeller & M.E. Halder), pp. 1541–1551. Amsterdam, The Netherlands: Elsevier Science B.V.
2. Dagg, A.I. (1999). Responsible animal-based research: three "Flags" to consider. *Journal of Applied Animal Welfare Science* **2**, 337–346.
3. Shapiro, K.J. (1998). *Animal Models of Human Psychology: Critique of Science, Ethics, and Policy*, 328pp. Göttingen, Germany: Hogrefe and Huber.
4. Singer, P. (1975). *Animal Liberation: A New Ethic for our Treatment of Animals*, 297pp. New York, USA: Avon.
5. American Medical Association (March 1988). *Use of Animals in Biomedical Research: The Challenge and Response.* Washington, DC, USA: Author.
6. LaFollette H. & Shanks, N. (1996). *Brute Science*, 286pp. London, UK: Routledge.
7. Dagg, A.I. (2000). Animal experimentation in cancer research: a citation analysis. *Journal of Applied Animal Welfare Science* **3**, 239–251.