2000

Evaluation of Animal Model Research

Kenneth J. Shapiro
Animals and Society Institute

Evaluation of animal model research

Published in Progress in M. Balls, A.M. van Zeller and M. E. Halder (Eds). *The reduction, refinement, and replacement of animal experimentation*, Elsevier 2000.

Kenneth J. Shapiro
Washington Grove, MD, USA

Abstract. It is argued that a concept of evaluation of animal models that is broader and more useful than validation is available. Productive generativity refers to the degree to which a model furthers understanding and leads to more-effective treatment interventions. Results of the application of this novel evaluative frame to several animal models of eating disorders show that this animal-based research has not been productive. The question of the relation between clinic and animal laboratory is discussed.

Introduction

How can we assess the value of animal models in medical and psychological research? This is primarily a problem of method — of developing rigorous evaluative tools. The focus in this paper is not on the question of the value of research, but on how to evaluate it. Obviously, an appropriate method must be open to finding any level of value, high or low.

We are concerned here only with science-based, not ethics-based, arguments regarding the use of animals in research. A science-based argument needs a scientific method to back its assertions in the current debate over animal research. Of course, all science-based arguments imply an ethical position. Among the several available philosophies on the issue — rights, utilitarian, social contract, feminist, communitarian — most science-based arguments assume some form of utilitarian ethic, in that they try to assess benefits to humans and costs to the non-human animals. The act of doing animal research is judged right or wrong exclusively by the test of usefulness.

For all but utilitarian philosophers, usefulness or value is, of course, a relative and limited consideration. To show that animal model research is useful is not necessarily to show it is justifiable. For example, for most people, invasive research on humans, although likely to be useful, is not justifiable. In addition, to show that animal model research was useful in the past, does not necessarily imply it was, or is, the most useful form of investigation. To show that some animal model research was useful, does not speak to the likelihood that most was not useful, and that some was harmful because it misled us.

With these constraints on the criterion of usefulness, the present paper addresses the problem of how to assess the utilitarian value of animal research. The focus is on animal model-based medical research (including psychological research). The paper describes a battery of available social scientific analytical methods supplemented by historical methods. The present author's background is in the social science side of psychology. For the sake of full disclosure, in ethics the author is a rightist. I hold that some animals other than humans, because they have inherent value, have basic rights which preclude their use in most medical and psychological research.

The present method is needed, because usefulness is not typically being assessed or assessed adequately in the scientific literature. Lafollette and Shanks end their theoretical analysis of the problem of assessment in their recent book, *Brute Science* [1], with a plea for further efforts to "assess the scientific value of animal experimentation. Until we do, we will be left with a vague version of the intellectually untenable 'it just works' argument." To contribute to the effort to "find ways to determine the contribution of animal experimentation to human well-being, especially in comparison to contributions derived from alternative research methods" is the intent of this present paper.

I will review some of the available literature on assessment, to extricate methods used, describe a battery of evaluative tools and show their application to a set of animal models, and will end with a few policy recommendations at the level of animal care committees. Along the way, I will suggest that validation is not a sufficient means of assessment of the value of animal research.

Literature review

I am spending significant time with earlier approaches, because they have not received much serious attention as methods and the present method has been described more fully elsewhere [2].

Some of the publications that purport to evaluate animal research are merely assertory, being not much more than lists of claims of benefits to humans or costs to animals. In my own field, Miller [3] provides a description of gains in effective treatment and in understanding of various psychological disorders that he claims derive from animal models. However, he provides virtually no evidence for those claims, and is frankly dismissive about costs to the animals. Miller notwithstanding, the simple presence of an animal model for a particular disorder itself provides no scientific evidence of the value or usefulness of the research based on it.

On the other side, Doncaster [4] and Diner [5] provide lists of procedures involving animals that are claimed by them to be costly to the animals, with virtually no discussion of possible benefits to humans. These studies are merely catalogues and, from the point of view of the present review of methods, are pre-methodological.

A number of authors have attempted to evaluate animal research by using historically based methods [6–9]). Paton distinguishes between the gains of increased knowledge and of applications. He also offers a historical method, but it is flawed and really begs the question of how to evaluate animal research. In his “method of deletion”, we would first identify what we have learned from animal research and then suppose that it is removed, in order to obtain a perspective on benefits. However, of course, identifying what we have learned is the crux of the problem. As I will discuss shortly, the literature in the history of science on the locus of discovery indicates that sorting out grossly between laboratory-based and clinic-based loci is difficult. Further, removing alleged benefits from animal research without replacing it with the benefits of whatever alternative scientific approach took its place is, even in the context of this imaginary exercise, illogical. For example, Drewett and Kani [10] argue that, in the field of psychology, cognitive-based therapies would have come earlier if animal-based behaviouristic psychology had not been so dominant.

The efforts of these scientists with historical bents have fostered some historically grounded discussion of the role of animal research in such controversial areas as the discovery of insulin treatment for diabetes, the development of understanding and prevention of polio, and the thalidomide tragedy. However, clearly we need to enlist the historians of science to enter the dense thicket of events surrounding these developments, discoveries, and mis-steps, and to help us to disentangle the claims and blames so often inflated by people on both sides of the issue.

Turning to early studies that are serious efforts at scientific methods of evaluation, one strategy employed can be called “most important advances.” Stephens analysed Nobel Prize winners in physiology and medicine [11]. He categorised the work occasioning the awards as animal-based or alternative-based, by using Russell and Burch’s concept for the latter category. He found that two-thirds of them used alternative methods. He also found that the relative percentage of alternative to animal research increased in recent years (up to

1985). Leader and Stark [12] also used Nobel Prize winners as an evaluative entry, but they failed to distinguish between animal and alternative methods.

In their 1975 study, Comroe and Dripps [13] employed an ambitious method to evaluate the origin of important gains in medicine. Grossly, they had specialists select the 10 most important advances in cardiac/pulmonary medicine out of a list of nominated advances. Then judges selected studies that contained the requisite body of knowledge for these advances. These studies were categorised as laboratory-based or clinic-based. The method has serious limitations from a social scientific point of view. For example, the bias of the judges is a serious problem, and one not simply resolved. In principle as well as in practice, it is difficult to locate judges that are both expert enough to select "essential bodies of knowledge" and yet are not either committed to laboratory or clinical approaches to research. Comroe and Dripps did not address this problem. However, in his study of the problem of elucidating locus of discovery, Reines [14] found a clear bias of clinicians to view the clinic and of basic scientists to view the laboratory, respectively, as the premier locus. He also found that claims of accidental discovery are inflated. More often, there is a considerable background of research which then leads to discovery. Teasing out the role of animals in that requisite background is difficult. Noting that the image of the serendipitous breakthrough of the lone scientist in the laboratory is a romanticised view, Reines takes the position that much discovery first occurs in the clinic, while later research in the laboratory has a merely demonstrative and graphic function.

The relatively new field of the sociology of knowledge offers general support for Reines' views. The description of science that emerges is of a process that is much messier than that provided in accounts by traditional philosophers of science. In place of a rationally driven process with strict rules of evidence and proof, Latour [15] and others describe a subculture of research where what occurs and what is credited as scientific facts (including those constituting significant advances) are largely intelligible as social constructions. It is a human enterprise where personalities, ambitions, and personal and political connections play significant roles.

In any case, the "most important advance" is an important advance over the pre-method approach of Miller and others. However, this approach really begs the question of how to evaluate the great preponderance of past, present, and proposed studies that clearly have produced or will produce no important advances. It is hardly arguable that the animal model strategy has not had some positive pay-off. Obviously, thousands of PhDs, over a period of 40 years, spending tens of billions of dollars (US) studying hundreds of millions of non-human animals in models of every conceivable medical and psychological disorder, produce some "important advances." What is at issue is not whether there have been at least some important advances in which animal research has played a role, but whether this is the most effective strategy. In addition to important advances, we need to consider wasted studies, wasted animals, mis-leads, and roads not taken through alternative and clinical methods.

Finally, I review here methods that employ the strategy of assessing both benefits and costs to evaluate animal model research. The Institute of Medical Ethics report describes features that should be taken into account in assessing *potential* benefits [16]. It includes social as well as purely scientific benefits. Under scientific concerns, they include quality of the proposed research — experimental design, necessity, and validity of procedures. This descriptive, but not yet quantitative scheme, is valuable, and the present method builds on it.

Kelly [17] and Giannelli [18], independently, in their responses to the Miller article mentioned earlier, use forms of citation analysis to evaluate animal model research. Giannelli notes the number of references cited by Miller as beneficial studies that appear in a popular licensure examination for psychologists. He finds the number to be small, at 5.9%. Kelly

examines the number of all citations in two high quality academic journals in one year that are reports of animal-based research, and finds low numbers, 0.003% and 2%. Both of these studies are methodological gains, but do not employ full citation analyses.

To conclude this review of the literature on evaluative methods, it suggests the importance of distinguishing between global, paradigm-specific, and study-specific evaluations. The pitfalls of assessment at the global level is that it tends to be merely assertory in making claims of benefits, it provides no middle range evaluation, and no way of assessing a particular study. This is unacceptable, because it is likely that most research contributes little, given the costs to the animals involved. Global approaches such as "most important advances" methods too readily lead to the sweeping premise that, since some animal research has been valuable, all of it is. This truncates the evaluation of any given or proposed study. Brody [19] offers this criticism in her analysis of the review system in the United States. We need methods that can evaluate any proposed study at the point of funding or animal review committee consideration.

Scope and terms of the present study

Before presenting a method consisting of a battery of analytical tools to that end, a few definitions are helpful. The method is intended for both medical and psychological animal model research, and we must be clear what this intended object of evaluation is and what it is not.

The distinction between medical and psychological research is a blurred one. Psychologists do physiology, and the field has several physiological subfields — physiological psychology, neuropsychology, and psychoneuroimmunology. Both enterprises have similar funding resources and use the same nomenclature for describing disorders. In the clinic, some psychological disorders have a physiological cause, some physiological disorders have a psychological cause, and all have psychological complications.

Another distinction that, upon closer examination, is blurred, is basic versus applied research. The intent of basic research is to add to the general understanding of biological processes, while the intent of applied research is to add to effective intervention (both preventative and direct treatment) regarding a particular human function or disordered function, and to understanding of that disorder. However, the effort to understand a particular disorder often involves investigation of general biological and psychological processes. In these terms, most animal model research is mixed, largely applied, but often, in part, basic also.

To concretise this, consider, as we will more fully below, an animal model of a human eating disorder. While the intent of much research with this model is to add to the understanding and treatment/prevention of bulimia or anorexia in humans, some research with the models aims at furthering understanding of eating, appetite, and satiation mechanisms at a fairly basic level of general biological function.

A similar argument about the blurring of applied and basic research apparently applies to toxicology research. While identification of toxins is a focus and has a direct applied end, toxicologists are also concerned with the mechanisms underlying, or the causes of, toxic reactions, and can pursue these at a basic level of research.

The fact of this blurring is important in the present context of evaluation of animal research, for there is some inclination to presume that basic research occupies a privileged region, that it is less accessible to evaluation, and therefore, is less accountable; or worse, that it is in principle less accountable, because it is basic.

Another term that needs clarification is "model." A model is an analogue. It is not the thing itself. It is similar to but not identical with what is being modelled. This is definitional,

but it is also arguable on theoretical grounds. LaFollette and Shanks [1] argue this for both evolutionary and systems theory, concluding that there are necessarily systemic disanalogies across species.

If it is only analogous, why do we use models? After all, one ideal of experimental science is the direct observation of the phenomenon of interest, not of some analogue of it. Why do we give up direct observation? The traditionally given reasons concern control, access, manipulation, and ethics. Another reason is that comparisons between two merely analogous phenomena helps us to think - it has a generative or heuristic function.

However, in response to the violation of the ideal of direct observation inherent in the animal model strategy, some maintain that animal models are not really that — that they do present the object under study for direct observation — forgetting that a disorder induced in an animal of a different species in a laboratory setting can never be more than more or less similar to that actual disorder in a human setting.

The proper function of a model is to help the scientist to think. We learn from differences as well as from similarities. Consider the animals of a species that, despite significant similarities, do not get a disorder that animals of a similar species do. Clearly, we might learn from comparisons between the model and the modelled.

Productive generativity refers to advances in understanding and advances in intervention, following Paton. Validation refers to the relative match between the model and the modelled. If the match is close, the model is validated. If it is not, or not in some critical ways, it is invalidated. Models can be quite similar to the modelled (a high-fidelity model) and thus validated, but still nothing new has been gained from the model. Conversely, the model can be quite different from the modelled (a low-fidelity model) and thus readily invalidated, but advances in understanding and/or intervention can still be occasioned by the research with the model. So a high-fidelity model can be low in productive generativity although readily validated, while a low-fidelity model can be high in productive generativity although readily invalidated. To reiterate, validation does not require that new knowledge or a new intervention has been discovered through the model. A validated model is not necessarily a productively generative one.

Method

To measure generative productivity (understanding and intervention), the method applies a battery of tests to evaluate the cost and benefits a body of research on eating disorders. The battery includes outcome study, citation analysis, survey, and a measure of pain, stress, and harm to animal subjects. These familiar quantitative methods are supplemented by informal historical inquiry. Outcome studies are sophisticated methods developed largely for programme evaluation. Typically, they involve several measures of an intervention before and at several points after its occurrence. There are several outcome studies of the effectiveness of the treatment of eating disorders in the literature (for example, Hsu [20]). Citation analysis is a measure of the frequency that published studies are mentioned in subsequent publications in the entire scientific literature. A refined version of citation analysis distinguishes varying qualities of the citation — is the cited study just mentioned in the review of literature, or were substantive or methodological findings influential in the citing study? Surveys are not as simple as they appear, and depend on the report of other people. When carefully done, they provide significant corroborating data.

There are several “pain scales” in the literature. I selected one developed specifically to rate the pain, distress, and harm done to animals in psychological research [21].

I applied this battery of measures to animal models of eating disorders, because it has been recognised, in recent decades, that they are major psychological disorders of near

epidemic proportion; the history of models developed and being developed goes back over two decades; the research does not have to be inherently painful or invasive (as in pain research); and, finally there have been no animal rights campaigns targeting researchers in this area.

The three animal models of eating disorders (ED) evaluated are the sham feeding and tail pinch models of bulimia, and the activity wheel model of anorexia. Bulimia involves recurrent episodes of bingeing and purging; self-starvation characterises anorexia. In both these ED, the client typically is a female adolescent or young adult who is preoccupied with her body-shape and weight. In sham feeding, the investigator surgically makes a hole in the wall of the stomach of a rat, dog, or non-human primate. This allows the investigator to feed the animal through the mouth and siphon off the food before it is fully digested. In this way, analogous to bingeing-purging, the individual "eats without calories". The tail pinch model features the importance of stress in bulimia, while the activity wheel mimics the role of exercise in the production and/or maintenance of anorexia.

Results

A review of treatment outcome studies shows that ED remain relatively intransigent to intervention. Treatments are only modestly and temporarily effective. Although there are some initial gains from treatment, these are limited to a reduction in the frequency of symptoms such as bingeing-purging, and relapse rates are high. To the limited degree that they are effective, the most common psychotherapy employed, cognitive-behavioural therapy, does not derive from animal models of these disorders.

While emphasising investigation of physiological mechanisms and the search for pharmacological treatments, the research involving these models has yielded no effective drug treatment to date. Fenfluramine, a drug earlier explored in research on obesity, was found to be ineffective in the treatment of bulimia, and was recently pulled off the market because it produced abnormalities in heart valves. Anti-depressant drugs are effective within the limits described, but they do not derive from these animal models.

As mentioned, citation analysis provides a measure of the frequency that published studies are mentioned in subsequent publications. Studies published on these three animal models by nine investigators were cited 0.69 times per year during the nine-year period considered (1986–1994). In comparison, the average annual frequency of all the references in the Science Citation Index is 1.87, or more than 2.5 times the rate of those examined in the present study. When only those citations judged significant are counted (cited only in a generalised introduction), the overall annual frequency drops to 0.31, i.e. 7 of 10 studies receive no significant citation in a given year.

A survey of clinicians specialising in the treatment of ED found that 60% did not know animal models of ED existed; 67% could not name or describe any model; 87% could not identify or describe the sham feeding model; and 87% indicated that animal models of ED did not influence their treatment approach. There was no overlap in the list of journals these specialists indicated they found "most helpful" in their work and those in which the studies of the nine investigators' were cited.

Application of the Invasiveness Scale found that these animal studies typically involve considerable pain, distress, and harm. The scale rates common experimental procedures on a 6-point scale, ranging from 0 to 5 (highest level). While it was developed to score psychological research, the scale correlates significantly with other, more-general pain scales. The models studied scored in the 3 and 4 range, which is consistent with the levels of invasiveness found in the field of psychology as a whole. For example, in the sham feeding model, animals are subjected to surgery, the distress of recovery, and harm and distress of a

permanent fistula which produces chronic indigestion. Further, many of the additional variables tested add to the level of invasiveness — brain lesioning, implanting electrodes into muscles involved in ingestion, and food deprivation.

Informal historical inquiry reveals that these models already existed in the laboratory as models of other disorders and as experimental procedures. They do not derive from direct observation of the clinical phenomena under study. For example, sham feeding is a procedure that has been used in the study of digestive physiology since the late 19th century [22]. The activity wheel was originally used as a model of ulcers. Tail-pinching has been used to induce stress, and was explored, at one time, as a model of schizophrenia.

The analogy between the model and the actual disorder is coarse, being based only on limited similarity. Sham feeding is only roughly analogous to binge-purge behaviour, as is the activity wheel to the inclination to over-exercise in anorexics. Tail-pinching produces stress, a generalised precursor to many, if not most, psychological disorders.

The ED animal model research emphasises physiological processes and the search for pharmacological treatment, although strong evidence points to a cultural basis of the disorders (the "slimming culture"). Finally, there is a preoccupation with developing technology and instrumentation. Investigators report on the development of metabolic chambers, tethering devices, wire implants, micro-lesioning, computer-based recording of neural and metabolic events, but provide little description of features of the disorders based on direct observation in their clinical settings.

A number of findings, taken together suggest that the animal model strategy, in practice, does not have enough intercourse with the clinic — it is ingrown. The source of models is based on already existing procedures in the laboratory, rather than on close observation and intimacy with the clinical phenomena; the relevant clinicians are unfamiliar with the models; there are limited validation efforts and these are largely internal to the laboratory enterprise (other models and relevant variables in the laboratory literature); and there are limited, if any, efforts to evaluate productive generativity.

Related applications and research

In the course of this study on ED models, I also informally examined models of aggression. There is a significant range of such models, which differ largely in the locus at which the induction of aggression occurs. Examples here are ordered from central to distal — drug induced (amphetamine, dopamine, anxiogenics), alcohol-induced, brain lesion (septal), pain, shock, conflict, conditioned suppression drinking conflict model, isolation (child abuse), frustration, cold, restraint, intruder. My impression is that these models are also low in productivity and, certainly, high in cost to the animals.

Dagg [23] used a method of flagging individual experiments on the basis of citation analysis, numbers of animals used, and invasive procedures. For example, a study with few citations, large numbers of animals used, and invasive procedures would receive three flags. By scoring large samples of these and noting university affiliation, journal published, and funding source, an investigator could compare universities, journals, and funding sources in terms of relative number of flagged experiments. Dagg scored 155 experiments by Canadian researchers published in 1990 in 14 journals devoted to animal research in psychology and neurophysiology. She found a large number of rarely cited experiments. On the basis of her work, she recommends that animal care committees and funding agencies use more-stringent criteria and require that applicants include a record of citations for each previously published study on related topics, particularly those using the same model or paradigm; she recommends pilot studies, particularly in studies involving large numbers of animals and highly invasive procedures.

A recent major survey of the attitudes of 4000 psychologists toward animal research, published in *American Psychologist*, found that 92% of psychologists who are primarily practitioners indicated that they rarely, never, or only occasionally used findings from psychological research on animals [24]. This result corroborates the results of my survey of specialists in ED. Incidentally, 95% indicated that a ban on animal research would not seriously hamper there work.

Recommendations

I have already reported recommendations from Dagg, emphasising relevant citation analyses and pilot studies for investigations at risk in terms of numbers of animals and high invasiveness

On the basis of my study, I have the following recommendations:

1) encourage independent investigators with expertise in using the above methods to assess the productive generativity of existing models;
2) animal review committees should alert researchers to the body of searchable literature produced by these enquiries;
3) animal researchers should conduct validation studies in the narrow sense, by formally testing the hypothesis of similarity between model and modelled; and
4) in conclusion, more broadly, the scientific community should be open to the possibility that the animal model research strategy is not the best available approach, on the grounds of both productive generativity and ethical considerations.

References

1. LaFollette H, Shanks, N. Brute Science. London: Routledge, 1996.
2. Shapiro KJ. Animal Models of Human Psychology: Critique of Science, Ethics, and Policy. Gottingen: Hogrefe and Huber, 1998.
3. Miller NE. The value of behavioral research on animals. American Psychologist 1985; 40: 423–440.
4. Doncaster A. Experiments on animals: A review of the scientific literature. Mississauga, Canada: Mississauga Animal Rights Society, 1982.
5. Diner J. Physical and mental suffering of experimental animals. Washington DC: Animal Welfare Institute, 1972.
6. Paton W. Man and Mouse: Animals in Medical Research. Oxford: Oxford University Press, 1984.
7. Kaufman S, Hahner K (eds). Perspectives on Medical Research. Volume 1. NY: Medical Research Modernization Committee, 1991.
8. Rowan A. Of mice, models, and men. Albany: State University of New York, 1984.
9. Reines BP. On the role of clinical anomaly in Harvey's discovery of the mechanism of the pulse. Perspectives in Biology and Medicine 1990;34:128–133.
10. Drewett R, Kani W. Animal experiments in the behavioural sciences. In: Sperlinger D. (ed) Animals in Research. New York: Wiley, 1981; 175–201.
11. Stephens M. The significance of alternative techniques in biomedical research: an analysis of Nobel Prize awards. In: Fox MW, Mickley LD (eds) Advances in Animal Welfare. Boston, MA: Martinus Nijof, 1987; 19–31.
12. Leader RW, Stark D. The importance of animals in biomedical research. Perspectives in Biology and Medicine 1987;30:470–486.
13. Comroe JH, Dripps RD. Scientific basis for the support of biomedical science. Science 1976;192:105–111.

14. Reines B. On the locus of medical discovery. The Journal of Medicine and Philosophy, 1981;16:183–209.
15. Latour B. Science in Action. Milton Keynes: Open University Press, 1987.
16. Smyth J, Boyd K (eds). Lives in the balance: The ethics of using animals in biomedical research. New York: Oxford University Press, 1991.
17. Kelly JA. Psychological research and the rights of animals: Disagreement with Miller. American Psychologist 1986;41:839–841.
18. Giannelli M. Three blind mice, see how they run: A critique of behavioral research with animals. In: Fox MW, Mickley LD (eds) Advances in Animal Welfare Science, 1985/86. Washington DC: Humane Society of the US, 1985; 109–164.
19. Brody M. Animal research: a call for legislative reform requiring ethical merit review. Harvard Environmental Law Review 1989;13:423–477.
20. Hsu L. Critique of follow-up studies. In: Halmi K (ed) Psychobiology and treatment of anorexia nervosa and bulimia nervosa. Washington DC: American Psychiatric Press, 1992; 125–150.
21. Shapiro K J, Field P. A new scale of invasiveness in animal experimentation. PSYeta Bulletin 1987;7:5–8.
22. Wolff H. Human gastric functioning: An experimental study of man and his stomach. New York: Oxford University Press, 1943.
23. Dagg AI. Responsible animal-based research: Three “Flags” to consider. Journal of Applied Animal Welfare Science, in press.
24. Plous S. Attitudes toward the use of animals in psychological research and education: Results from a national survey of psychologists. American Psychologist 1996;51:1167–1180.

Footnote to title page

Address for correspondence: Dr Kenneth J. Shapiro, P.O. Box 1297, Washington Grove, MD 20880, USA.