Analogical reasoning with animal models in biomedical research.

Cameron Shelley
12 Clark St. W.
Guelph, ON
Canada N1H 1S9

Introduction

What makes animal models valid? In other words, what assures us that experiments performed on animals can tell us *anything* about conditions in human beings? In addition to being an interesting question in the philosophy of science about a crucial scientific research practice, the question of the validity of animal models looms large in the continuing debates concerning animal rights. One sort of argument for the abandonment of animal modeling attacks its validity as a scientific research practice. The general idea behind this sort of argument is that if animal modeling can be shown to be non-valid, then it is obviously morally unacceptable. After all, if we cannot show how claims about human conditions can be tested by experiments on animals, then those experiments ought not to be undertaken at all. So, the validity or non-validity of animal models becomes a pressing question for a variety of reasons.

Biomedical researchers generally agree that the validity of animal models depends significantly on, among other things, the existence of *analogies* between animal models and their human targets. In short, the validity of an animal model depends on whether it is analogous to the human condition it is supposed to represent. In their recent book *Brute Science*, philosophers LaFollette and Shanks (1996) have attacked the practice of animal modeling by attacking this source of its validity. Specifically, LaFollette and Shanks argue that *even when* models are analogous to their targets, there remain *disanalogies* between model and target that undercut *any* such claims for the validity of an animal model. So, they conclude, animal modeling is also an immoral practice.

My first goal in this talk is to defend the validity of animal models from LaFollette and Shanks's critique. I will argue first that LaFollette and Shanks's attack is ineffective because they construe the concepts of analogy and of disanalogy in a way that is largely irrelevant to the validity of animal models. However, I will also argue that if we *re-construe* these concepts appropriately, then we can construct an argument that is potentially devastating to the validity of animal models. This reconstructed argument involves employing the concepts of analogy and disanalogy developed in recent cognitive science research (Shelley 2002a, 2002b). Then, I will explain why I think that even the reconstructed argument does not establish the general non-validity of animal modeling.

So, in the first instance, I aim to rescue the validity of animal models from LaFollette and Shanks's attack. However, what I ultimately aim to accomplish is to show how recent research on analogical reasoning in cognitive science can tell us, with far greater clarity than before, what makes animal models valid.

Analogy and face validity

As I mentioned earlier, LaFollette and Shanks argue for the abandonment of animal models on the grounds that the practice is irretrievably non-valid. In this part of the paper, I will present LaFollette and Shanks's argument as they give it, and point out why the argument falls short of establishing the non-validity of animal models. My criticisms of this argument will help to motivate my reconstruction of it afterwards.

LaFollette and Shanks's argument can be very briefly summarized. It begins with the concession that animal models derive a kind of *prima facie* validity from relevant analogies between animal model and human target. Rats and humans, for example, are indeed alike in many ways; so we might reasonably conclude that it makes sense to compare them. However, the argument continues, there are always disanalogies between animal models and human targets, and these disanalogies have the effect of irretrievably undermining the validity of animal-human comparisons. Rats and humans are, after all, unlike in many ways. So, animal models are, in general, scientifically non-valid, and therefore immoral.

There are a number of flaws in LaFollette and Shanks's argument. A crucial flaw concerns the concepts of analogy and disanalogy that they apply. Specifically, LaFollette and Shanks understand analogy to be the similarity or *resemblance* between model and target. This understanding becomes evident when they begin their argument by defining an analogical inference as a sort of syllogism in the tradition of J. S. Mill (1872):

> X has properties *a,…,e, f.* (X = model, from which information is drawn)
> Y has properties *a,…,e.* (Y = target, to which information is to be added)
> Probably, Y has property *f.*

In other words, two things X and Y are analogous insofar as they have the same properties, such as color, shape, mass, or constitution. Since some resemblance between X and Y has been established, it is probable that X and Y share some further property *f* in common as well. L&S then go on to discuss animal models in terms of what they call a *Causal Analog Model* (CAM). They define a CAM by adding three extra constraints on the definition just given. These constraints are:

1. the common properties *a,…,e* must be causal properties that
2. are causally connected with the property *f* we wish to project, and
3. there must be no causally relevant disanalogies between X and Y.

The term *causal* simply means that the properties *a,…,e* give rise to each other through some means. More importantly, note how the third constraint does the damage to animal modeling. It asserts that the presence of *any* causally relevant disanalogy completely undermines the comparison between model and target. Now, LaFollette and Shanks do not define what they mean by disanalogy, but the obvious and conventional meaning in the same tradition is that a disanalogy is a *dissimilarity*. In other words, a model X and a target Y are disanalogous if there is some property, call it *p*, possessed by either X or Y but not by both. Of course, there are almost countless dissimilarities between human beings and any given variety of animal, some of which are sure to be causally relevant. Given this third constraint on animal modeling, this fact means that any animal model is sure to be non-valid.

Let us accept for the sake of argument that there will always be causally relevant dissimilarities between animal models and human conditions. Does this claim truly imply that animal models are always non-valid? The answer, I think, is *no*. Very little

depends upon the resemblance as such between model and target. In the biomedical literature, the validity that an animal model derives from its similarity to its target is known as *face validity* (see Willner 1991). For example, a rat that moves less than usual for its kind, or that consumes less food than usual for its kind might be considered as a model of depression in humans (Sarter and Bruno. 2002). These properties, less movement and less food consumption than typical, are characteristic symptoms of depression in people. Because of these shared properties, this rat model of human depression enjoys some face validity. However, rats, as Sarter and Burno (2002) point out, are not known to write poetry about their experiences in life. Depressed humans are known to do this, so a rat model of human depression would seem to be lacking somewhat in face validity.

However, face validity counts for very little in the evaluation of animal models. As biomedical researchers are well aware, dissimilarities between species are often just the result of differences in the behavioral repertoires of the animals concerned. Two sorts of behavior, for example, may be analogous even though they are dissimilar in causally relevant ways. For instance, giving stimulants to rats induces a behavior of stereotyped rearing on the hind legs. In contrast, giving stimulants to primates induces a behavior of stereotyped scratching (Willner 1991, 14). Now, rearing and scratching are dissimilar in the sense that they do not resemble one another to the eye. In other words, they are easily told apart. Nevertheless, they are analogous in the sense that both behaviors are the results of treatment with stimulants.

So, the central issue when evaluating comparisons between animals and animals, or between animals and humans, is whether or not the conditions being compared are functionally equivalent. Stereotyped rearing and stereotyped scratching are functionally equivalent as responses to stimulants.

It is because of considerations like these ones that biomedical researchers set little store in the face validity of animal models (Davidson et al. 1987). This being the case, LaFollette and Shanks's critique of animal models concerning their face validity is of little consequence. However, we now have a concern, namely functional equivalence, that indeed a crucial consideration.

Analogy and construct validity

We now have good reasons to reject LaFollette and Shanks's argument against the validity of animal models in general. However, we now have an issue, namely functional equivalence, that is far more important to the validity of animal models. It is now time to show how recent research on analogical reasoning in cognitive science can help us to clarify what makes animal models valid. I will then be in a position to reconstruct LaFollette and Shanks's argument in a way that makes it far more penetrating.

When considering functional equivalence, biomedical researchers refer to the *construct validity* of an animal model. In terms of animal models, construct validity concerns how well the condition of a model *represents* the condition of interest in the target (Willner 1991). In other words, a model must correspond to its target in some appropriate way. In general, biomedical researchers have had no explicit way of conceptualizing and assessing this form of correspondence. They simply rely on their intuitions. From the preceding discussion, it is clear that similarity is not the right way to think about this correspondence. Instead of capturing correspondences in terms of shared

properties, we need to capture correspondences of equivalent functions. I claim that the concept of analogy as it has developed recently in the cognitive science literature captures exactly this idea. That is to say, cognitive scientists have come to view analogies as a particular sort of functional correspondence and have supplied tests for the assessment of analogies. So, I think that we can make substantial progress in clarifying and assessing the construct validity of animal models by applying this work on analogies.

Consider, for example, a common mouse model for the action of antidepressants in human beings. In the Porsolt Forced-Swim Test (Porsolt et al. 1977), a mouse is placed in a cylinder of water and watched to see how long it swims until it gives up trying to climb out. (The mouse is not drowned at that point but assumes a static position with its hind feet on the cylinder bottom and its nose out of the water.) It turns out that mice that are treated with antidepressants tend to swim longer than normal mice do. This model of the action of antidepressants enjoys construct validity in the sense that the increased time that mice treated with antidepressants spend trying to extricate themselves from the cylinder corresponds to the increased hope for success in life that depressed people treated with antidepressants feel in the pursuit of their goals. The analogy between mouse and human that is present in this case may be captured as in the table below.

Mouse	**Human**
mouse	patient
antidepressant	antidepressant
safety	goals
longer-time	further-extent
look-for(mouse,safety)	hope-for(patient,goals)
persist(mouse,longer-time)	persist(patient,further-extent)
receive(mouse,antidepressant)	receive(patient,antidepressant)
because(persist,look-for&receive)	because(persist,hope-for&receive)

This table represents the analogy in accord with the principles of the *structure-mapping theory* (Gentner 1983). The information in the mouse-human analogy is laid out according to a few simple principles. All the information concerning the model domain, that is, the mouse, is located in the left-hand column. All the information concerning the target domain, that is, the depressed human patient, is located in the right-hand column. Each row in the table represents a correspondence between the model and target. For example, the mouse and the human patient are both placed in the same row (at the top) because the mouse and patient correspond to each other in the analogy. Now, there are three different kinds of correspondences that may be found in an analogy, and these are captured as three groups of rows in the table. The topmost area in the table, consisting of the top four rows, represents correspondences between the simple properties or *attributes* of each domain, such as *safety* and *goals* in row three. The middle area, consisting of the next three rows, represents correspondences between *relations* among the attributes or facts regarding the attributes, such as *look-for(mouse,safety)* □ *hope-for(patient,goals)* in row five. In plain English, what this row states is that the search for safety by the mouse corresponds to the hope for achievement of goals by the patient. The bottom area, consisting of the last row, represents correspondences between the *system relations* of each domain, such as *because(persist,look-for&receive)* □ *because(persist,hope-for&receive)*. This row states that just as the persistence of the test mouse for a longer

time is due to its search for safety and its treatment with antidepressants, the persistence of the psychiatric patient to a further extent is due to his or her hope for achievement of life goals and treatment with antidepressants.

We can see from looking at the representation in the table that the content of the analogy is unified by the system relations. The *because* relations, in this case, summarize the whole analogy when rendered in plain English. A look at the table shows that all the basic attributes and relations in each domain are related to each other by falling under the system relations. It is these high-level relations that provide the overall structure of the analogy and thereby determine the mutual relevance of each piece of information contained within it.

This way of representing the corresponding relations in an analogy, anchored by the system relations, captures the functional equivalence between model and target. In doing so, it captures the construct validity of the model. So, we can evaluate the construct validity of this animal model by evaluating this analogy. The multiconstraint theory of analogy provides three criteria for performing this evaluation (Holyoak & Thagard 1995). The better the analogy satisfies these constraints, the stronger or more coherent the analogy is considered.

1. Structural consistency: This constraint concerns to what extent the analogy presents a structural isomorphism. Looking at the table, we can see that every term in each domain corresponds to exactly one and the same term in the other domain. For example, *mouse* and *patient* correspond at every point where those terms occur. Since every correspondence is one-to-one in this way, the analogy is completely consistent structurally.
2. Semantic similarity: This constraint concerns to what extent corresponding terms are similar in meaning, particularly at the relation and system-relation levels. For example, *look-for* and *hope-for* are similar although not identical, and *because* (mouse) and *because* (patient) are identical. Since the corresponding terms at these levels are similar or identical in meaning, the analogy is strong in this respect also.
3. Pragmatic coherence: This constraint concerns how well the overall story suggested by the analogy addresses the facts at hand. In this case, the analogy suggests how treatment with an antidepressant relates to subsequent behaviors in both mouse and human. Since it helps to make sense of these observations, the analogy can be considered pragmatically coherent as well.

Since these constraints are well satisfied by the mouse-human analogy, the analogy is clearly very coherent overall. In turn, its coherence implies that this mouse model of antidepressants in human beings enjoys strong construct validity: the longer safety-seeking of the mouse in the model well represents the greater hope for achievement of goals in depressed human beings.

Disanalogies and construct validity

I have now illustrated how construct validity can be captured by the concept of analogy developed in cognitive science research, such as the multiconstraint theory. I can now raise an objection similar to LaFollette and Shanks's argument concerning face validity. In brief, what happens to our assessment of the construct validity of animal models when there are disanalogies between model and target?

LaFollette and Shanks showed that models that are causally similar to their targets may also be causally dissimilar as well. This observation tends to undermine the face validity of animal models. Now, face validity is a matter of relatively little concern, but what if a similar problem applies to the *construct validity* of animal models? Perhaps models that are analogous to their targets in terms of the multiconstraint theory may also be disanalogous as well. If true, this situation means that the construct validity of any animal model, such as the Forced-Swim Test is not assured by the existence of even a strong analogy between model and target. If so, then the validity of animal models in general is cast into doubt.

In order to explore and assess this argument, we need to find an appropriate concept of disanalogy. The concept of disanalogy that I have developed in my recent work fits exactly this requirement. In Shelley (2002a, 2002b), I defined a disanalogy as an extension to a given analogy that supports a conclusion incoherent with the conclusion supported by the original. In other words, a disanalogy arises when we have an analogy that leads us to one conclusion initially but that leads us to an incompatible conclusion when augmented with new information.

To clarify this concept, let us consider an example of a disanalogy to the Forced-Swim Test (Schatzberg, in Kolata 2004). In a mouse, the effect of an antidepressant is most likely produced by affecting the lower areas of the brain. The frontal cortex in a mouse is small in relation to its lower brain regions in contrast with the relatively large frontal cortex of a human being. So, since the antidepressant is probably acting on a brain area *associated* with non-cognitive mental states, it is likely that it is affecting the mouse's non-cognitive mental state. Now, we can add this information to the existing analogy between mouse and human and draw the corresponding conclusion regarding the human target. That conclusion is that the mental state induced in a depressed person by treatment with antidepressants is also a non-cognitive mental state.

Here we encounter some incoherence. This conclusion does not fit with the rest of the analogy. In particular, the hope that a depressed person experiences as a result of treatment, represented by the *hope-for* relation in the analogy, is an irreducibly cognitive state. People do not merely hope, they hope for some things in particular, like publishing a paper or being promoted at work. People represent these hopes to themselves cognitively, in the form of mental representations, using their big frontal cortices. Yet, from what we were just saying, the hope for achievement of life goals in these patients should be like the safety-seeking in mice: instinctual rather than involving sophisticated mental states. It appears very much as though this one mouse model of human depression is leading us to two, incompatible conclusions. If this is so, then we should probably consider the whole comparison misleading and so reject the construct validity of the Forced-Swim Test.

This disanalogy is represented explicitly in the table below. This table is simply a copy of the previous one with the new information added. The added correspondences are indicated with asterisks. The new causal relation in the bottom row states that the effect that an antidepressant has on a mouse's non-cognitive state is what causes the mouse to persist for a longer time when looking for safety. The contents of this causal relation are also added in as the *affect* relation and the *non-cognitive-state* attribute in the appropriate places. Equivalent representations have been placed in the human target

domain. The result is a new analogy that appears to be as coherent as the original one in terms of the multiconstraint theory.

Mouse	**Human**
mouse	patient
antidepressant	antidepressant
safety	goals
longer-time	further-extent
*non-cognitive-state	*non-cognitive-state
look-for(mouse,safety)	hope-for(patient,goals)
persist(mouse,longer-time)	persist(patient,further-extent)
receive(mouse,antidepressant)	receive(patient,antidepressant)
*affect(antidepressant,non-cognitive-state)	*affect(antidepressant,non-cognitive-state)
because(persist,look-for&receive)	because(persist,hope-for&receive)
*cause(affect,persist)	*cause(affect,persist)

It is clear that this new analogy meets the definition of a disanalogy given above. It is derived from the existing mouse-human analogy by adding new information to it. Furthermore, the new information suggests a conclusion: that the antidepressant affects a non-cognitive mental state. This conclusion does not fit with the implications of the original version of the analogy. Therein lies the problem: Essentially one-and-the-same analogy suggests two, mutually incompatible conclusions. Given this situation, it seems that we cannot rely on *anything* that this analogy suggests.

Defeasibility and construct validity

I have now given a reconstructed version of LaFollette and Shanks's argument aimed at construct validity. As things stand, the whole idea that we can sensibly compare mice and depressed humans is in jeopardy. When we generalize this difficulty, we get an argument threatening the construct validity of any animal model. So, it seems that the validity, and therefore morality, of the whole practice has been thrown into doubt. What I will now argue is that this argument to the general non-validity of animal models does not work either. The reason is that disanalogies, as I have described them, are *defeasible*.

In general, to say that an inference is defeasible is to say that its conclusion may be retracted in the light of further evidence. In the case of disanalogies, defeasibility means that the incompatible conclusions suggested by the disanalogy may be reconciled with each other by the introduction of new information into the context of the analogy. For example, one might point out that cognitive and non-cognitive states are not independent in humans: The frontal cortex *interacts* with lower brain areas in the realization of mental states. So, even if antidepressants do act in the first instance on the non-cognitive component of mental states in mice and humans, nothing prevents them from thereby affecting the *cognitive* states of humans as well. Indeed, such an interconnection is to be expected given the strong emotional component of depression, which is strongly associated with non-cognitive areas of the brain. So, the two conclusions supported by the disanalogy discussed earlier seem to fit together after all: It is not so odd to think that antidepressants might act on both cognitive and non-cognitive mental states at the same time.

This example illustrates why the general argument against the construct validity of animal models does not work. Whether or not a disanalogy succeeds in undermining the validity of an animal model depends upon the details of the particular model in question. In some cases, there may be no way of reconciling the incoherence produced by raising a disanalogy. In other cases, such as the one we have discussed here, there are reasonable ways of resolving that incoherence. Each case will depend upon the context of each model. So, no general conclusion about the construct validity of animal models follows from these considerations. What we can say in general is that, with the proper concepts of analogy and of disanalogy in hand, we can hope to make reasonable determinations of construct validity in each individual case.

Conclusions

The most immediate conclusion that we can draw from this discussion is that the argument made by LaFollette and Shanks against the validity of animal modeling in general is not effective. In its original form, it addresses face validity, which is a relatively inconsequential matter. When revised to address construct validity, the argument becomes potentially more damaging. However, because the argument relies on disanalogies and because disanalogies are defeasible, no general conclusions follow about the construct validity of animal models.

More importantly, I argue that I have greatly clarified one important answer to the question in the philosophy of science of what makes animal models valid. One vital quality is the construct validity of animal models. The best way to understand the construct validity of animal models, I claim, is to apply recent cognitive theories of analogy—and of disanalogy. I have tried, in particular, to show how consideration of disanalogies is indispensable to understanding the validity of animal models.

Finally, these theories help us not only to better understand what makes an animal model valid, but they also provide us with the means for principled evaluation of them. Having these conceptual tools is extremely important, given the continued reliance on animal models in biomedical research, and given the ethical controversy regarding them that continues today.

Acknowledgments

Previous versions of this paper were given as talks at the University of Waterloo and at Carleton University. Thanks to Lorenzo Magnani for making it possible for me to give this talk at the Model-Based Reasoning conference in Pavia in December 2004.

References

Davidson, M. K., J. R. Lindsey, and J. K. Davis (1987) Requirements and selection of an animal model. *Israel Journal of Medical Sciences 23*: 551–555.

Gentner, D (1983) Structure-mapping: A theoretical framework for analogy, *Cognitive Science 7*(2): 155–170.

Holyoak, K. J. and P. Thagard (1995) *Mental leaps: Analogy in creative thought*. Cambridge, MA: MIT Press.

Kolata, G. (28 March 2004) Why test animals to cure human depression? *New York Times*.

LaFollette, H. and N. Shanks (1996) *Brute Science: Dilemmas of animal experimentation*. London: Routledge.

Mill, J. S. (1872) *A system of logic*, 8th edition. London: Longman.

Porsolt , R. D., A. Bertin, M. Jalfre (1977) Behavioral despair in mice: a primary screening test for antidepressants, *Archives internationales de pharmacodynamie et de therapie* 9(2):327–36.

Sarter, M. and J. P. Bruno (2002) Animal models in biological psychiartry. In H. D'Haenen, J. A. den Boer, and P. Willner (eds.), *Biological Psychiatry*, pp. 1–8. Hoboken, New Jersey: John Wily & Sons.

Shelley, C. (2002a) Analogy counterarguments and the acceptability of analogical hypotheses, *British Journal for the Philosophy of Science 53*(4): 477–496.

Shelley, C. (2002b) The analogy theory of disanalogy: When conclusions collide, *Metaphor and Symbol 17*(2): 81–97.

Willner, P. (1991) Methods for assessing the validity of animal models of human psychopathology. In A. A. Boulton, G. B. Baker, and M. T. Martin-Iverson (eds.), *Animal models in psychiatry*, pp. 1–23. Clifton, New Jersey: Humana Press.