Stafford, T. "What use are computational models of cognitive processes?"
In Mayor, J., Ruh, N., Plunkett, K. (2009). Connectionist Models of Behaviour and Cognition II: Proceedings of the 11th Neural Computation and Psychology Workshop. World Scientific.

# What use are computational models of cognitive processes?

Tom Stafford
Department of Psychology, University of Sheffield

Computational modellers are not always explicit about their motivations for constructing models, nor are they always explicit about the theoretical implications of their models once constructed. Perhaps in part due to this, models have been criticised as "black-box" exercises which can play little or no role in scientific explanation. This paper argues that models are useful, and that the motivations for constructing computational models can be made clear by considering the roles that tautologies can play in the development of explanatory theories. From this, additionally, I propose that although there are diverse benefits of model building, only one class of benefits — those which relate to explanation — can provide justification for the activity.

## What use are models?

What kind of object are computational models, and how are they scientifically useful? Among the modelling community there is little in depth discussion of these issues. This is partly, we may suppose, because among the converted there is little need to rehearse doctrine. But even in textbooks the philosophical status of modelling *per se* takes second place to details of specific models and some introductory discussion of specific issues such as level of representation. (Ellis & Humphreys, 1999; Elman, 1996; O'Reilly & Munakata, 2000) This can give the impression that the nature of modelling with regard to scientific explanation is well understood. As modellers we *know* that models are useful; indeed, our work is based on this assumption. However, not everyone shares this feeling, nor agrees with this position. In fact there have been sustained debates over the proper use and purposes of modelling. (Lewandowsky, 1993; Roberts & Pashler, 2000; Smolensky, 1988)

One point of contention, which this article will use as a starting point, is that computational models (henceforth 'models') are defined in mathematical terms and so can appear to share with mathematics the property of being tautological. This has led some to suggest that models cannot tell us anything new about the world. (Segalowitz & Bernstein, 1997). Models, it is claimed, can make predictions but this is their only role in the scientific process. They cannot ever be part of a "test of how humans actually work" nor can they "provide new information about brain organisation or function" (Segalowitz & Bernstein, 1997). An important related idea is that computational models can be uninterpretable black boxes — a-theoretical objects which may match human performance or structure but which do not provide any additional information, precisely because a model could have been built which would match any pattern of data, not just this specific one. (McCloskey, 1991). The claim is that, because the workings of a model are unintepretable or irrelevant to psychology or neuroscience, their only use is to make predictions which *can* be compared to psychology or neuroscience. Models, in this view, are in no way analogous to real theories of psychology or neuroscience (which are verbally or logically defined).

Beyond this, there has also been criticism of the 'glamour' of computational modelling. Nobel laureate Francis Crick was extremely sceptical of the early Parallel Distributed Processing (D. Rumelhart, McClelland, & the PDP Research Group, 1986) movement:

> "I also suspect that within most modellers a frustrated mathematician is trying to unfold his wings. It is not enough to make something that works. How much better if it can be shown to embody some powerful general principle for handling information, expressible in a deep mathematical form, if only to give an air of intellectual respectability to an otherwise rather low-brow enterprise." (Crick, 1989)

The implication is that modelling persists in recruiting practitioners and advocates because it has an air of mathematical rigour and complexity, while actually this disguises what is no more than a kind of grubby *ad-hoc*ism — some spur-of-the-moment engineering solutions motivated merely by fitting the data. And, worse than this, Crick suggests, modelling is a particularly non-informative kind of *ad-hoc*ism.

Clearly the utility of modelling is not as well established as you might suppose from attending a gathering of modellers, such as one of the Neural Computation and Psychology Workshops, or from reading a proceedings such as this one.

This article is underpinned by the belief that it is a lack of understanding of the nature of models and modelling that

Thanks to Hubert Petre for the loan of his flat in Brussels while I considered these issues, to Stuart Wilson for helpful comments on a draft of this article and to the students of the Computational and Cognitive Neuroscience Masters, University of Sheffield, for several useful discussions.

leads to confusion over their scientific value. Since models undeniably involve mathematical tautologies, and since this tautological nature has been leveled as a criticism against modelling, I will use the analogy of a tautology to explore the possible benefits of doing computational modelling. Through this I hope to clarify what theoretical work modelling can hope to do.

## A very simple tautology

The code that runs a model can be expressed in mathematical equations, and it is mathematical element that makes a model tautological. However, the heart of any model is the effort to establish a correspondence between parts of the system being modelled and the parts of the model. As Kenneth Craik, a prescient theorist of cognitive science, said

> "By model we thus mean any physical or chemical structure system which has a similar relation-structure to that of the processes it imitates" (Craik, 1943)

Models will necessarily be tautological with respect to their component parts. Because it is possible to reduce a model to a set of mathematical equations this must be true. But our scientific interest in a model lies not merely in the equations as such, but in the relation of the component parts of the model to component parts of the world. Models are formally specified by their equations, but they are also comprised of a set of model-world relations. Because of this they are more than "black-boxes" and so can inform our theories of the world in deeper and more complex ways than merely making predictions.

As a vehicle to explore the different ways in which tautology can, in fact, inform us about the world, let us begin by taking a very simple tautology: that 1 + 2 = 3. Model simulations only differ from this tautology in their degree of complexity, and their consequent opacity. This article will hope to use the very simplicity of this particular tautology to illustrate the value of models-as-tautologies in general.

## The importance of tautology

So, let us begin to consider the possible ways in which a tautology can inform a scientific theory.

### *Sufficiency*

The first way is the demonstration of sufficiency. Imagine that '3' in 1 + 2 = 3 is a known real-world phenomenon. The model can demonstrate that other phenomena ('1' and '2') and known causal laws ('+') are sufficient to produce it.

Whether or not this is interesting theoretical work depends on the current beliefs of theorists about how '3' arises. When the existence of the result ('3') is uncontroversial but the ingredients are uncertain, we have provided a theoretical option, a possibility - the status of which depends on research into whether the ingredients exist, and on the number of other candidate theories there are for producing this result. An example of this kind of work is Linsker's (1988) demonstration that self-organisation according to the Hebb rule can produce cells with receptive field properties like those found in the visual system. This model does not tell us whether the modelled dynamics actually are responsible for the receptive field properties of visual system neurons. It merely demonstrates rigourously that this possibility exists, and consequently, raises our assessment of the likelihood of this hypothesis being true.

### *Prediction*

If the presence of the ingredient elements ('1 + 2') is uncontroversial but the result ('3') is either not known or not commonly associated with these ingredients, then the model makes a prediction: that the result element will arise from the ingredients. An example of this is McClelland & Rumelhart's (McClelland & Rumelhart, 1981) predictions concerning the effect of word context on letter recognition. These predictions were experimentally confirmed by Rumelhart & McClelland. (D. E. Rumelhart & McClelland, 1982) This ability to make predictions which can be confirmed or falsified is obviously a core part of the scientific process (Popper, 1968), and is often offered as a major motivation for building models. The issue is discussed further below in relation to the desirability of cumulative modelling programmes.

### *Existence proof*

Even if all the elements in the models are uncontroversial, then modelling can still provide an informative result by establishing a possible connection between the ingredients and the results. This is a variety of what is known as an existence proof. An example is Plaut & Shallice's (Plaut & Shallice, 1993) demonstration that attractor dynamics in the orthography-semantics mapping can produce the pattern of errors found in patients with deep dyslexia. The existence of attractor dynamics is not controversial, nor is the pattern of errors found in deep dyslexics. What the model established was that attractor dynamics could be the source of the pattern of errors, a possibility that was hitherto not regarded as a plausible hypothesis.

An existence proof style model does nor prove what *is* happening, merely what *could be* happening. Once an existence proof is offered the processes which do in fact cause some result still need to be investigated. How scientifically interesting an existence proof is depends on the current opinion about the mechanism illustrated. If it is controversial or novel the model is obviously more interesting.

### *Insufficiency*

Equally useful, but less common than the purposes discussed above, is the demonstration of insufficiency. Consider again our tautology '1 + 2 = 3'. An implication of this tautology is that '1 + 2' equals 3 and no more than 3. In the case that some other result ('4') is known to exist, the model demonstrates the insufficiency of the hypothesized ingredients to produce it, and thus provokes a search for the additional factor which must be present, given that the ingredient

elements have been shown to be insufficient alone. An apocryphal example (perhaps arising from Hurlbert & Poggio, 1988) is the story told of AI pioneer Marvin Minksy assigning 'vision' to a graduate student as a summer project. The point of the story is to illustrate how mistaken the scientific community was about the difficulty of vision as a problem. It was through a generation of researchers in AI attempting to build models which could recognise what they saw, and thus discovering that all known methods were insufficient to achieve this, that it was realised how hard the problem of vision really is. The demonstration of insufficiency is also important in establishing what results would contradict or falsify a model. (Roberts & Pashler, 2000).

## Models as theories, theories as explanations

The above framework should make clear that the importance of models resides not in their formal structure alone, but in their purported correspondence to certain features of the world. The usefulness of a model lies in how it informs us about the potential relationships between features of the world.

An informal survey of modelling work presented at the 11th Neural Computation and Psychology Workshop (Oxford, July 2008) suggests that the most common purpose for which models are constructed — or at least the most common justification offered for their construction which falls within the current framework — is that of sufficiency: the demonstration that a certain set of ingredients is capable of producing a certain outcome. A danger of this kind of work is that, as previously noted, models may, if insufficiently constrained, match any possible outcome. Roberts & Pashler (Roberts & Pashler, 2000) have argued convincingly that the value of a model can only truly be assessed when that data which is *cannot* fit is also made clear. This relates to the idea of insufficiency in the current article.

This framework also helps draw out why realism *per se* is not the only metric on which models should be compared. The virtue of a model lies not in the number of biological details it contains, as such, but rather in its accuracy of correspondence with phenomenon at the level of description that it is trying to model. Indeed, one core motivation for modelling in the first place is to develop useful abstractions. Al Wilhite (personal communication) made this point eloquently by modifying a quote from Guy de Maupassant's essay 'The Novel', replacing the word 'artist' with that of 'theorist':

> "The [theorist] will endeavour not to show us a commonplace photograph of life, but to give us a presentment of it which shall be more complete, more striking, more cogent than reality itself. To tell everything is out of the question"

It is for this reason that more details are not necessarily better; sometimes models can be improved by including less detail rather than more. This point is a fundamental one when considering the choice between competing theories, going back to William of Ockham (c. 1288 - c. 1348), through to modern information theoretic formulations of criteria for model comparison (for an introduction, see chapter 28 of MacKay, 2003). Nonetheless it is a point that still needs to be made concerning the modelling of cognition (Dror & Gallogly, 1999). Obviously readers will have their own beliefs about the difficulties and benefits of modelling at different levels of description. There is a strong case to make that since one great strength of computational modelling is to connect psychological and neuroscientific levels of explanation, more biological detail will frequently yield models with greater explanatory powers, but this debate is beyond the scope of the current article.

If we accept that models are more than their mathematical constitution and are also comprised of assertions about the correspondence between their parts and the world, then we must acknowledge that models are more than mere black-boxes. In fact, models are theoretical entities, albeit with more formal constraints and distractions that verbally specified theories, and with the proviso that models are underdetermined by theories (analogously to the way that theories are underdetermined by data — any theory will have a family of modelling implementations). By making connections between known and proposed entities, models do the work of theories. If this is accepted then, like theories, models provide explanations. Furthermore their explanation-providing capacity is inextricably linked to their tautological nature.

There are deep and sustained issues concerning the philosophy of explanation. (Mayes, 2008) The framework above is an attempt to illustrate the important theoretical work that can be done by models as tautologies. A similar, but far more thorough, exposition of the importance of this kind of work has been done by Kulka (2001; 1995) in relation to non-computational theories in psychology, using the terms 'theory amplification' (discovering necessary predictions, inconsistencies, complementarities and postdictions between theories and data) and 'simplification' (the application of different kinds of parsimony).

## A proposal

There are other benefits of modelling beyond this framework suggested by the consideration of models as tautologies. Modelling allows us to work with quantitative and multicausitive theories beyond the vague intimations of descriptive theories. Models can help us define a problem (Marr, 1982) or allow us to integrate different, even conflicting, theories at different levels of description within the same framework. Finally, models have considerable value for cultivating the intuitions of individual researchers. As Paul Krugman said, "We just don't see what we can't formalize". (Krugman, 1993) For many modellers a primary value of modelling is the challenge to and cultivation of their intuitions that they make possible.

Notwithstanding these other benefits, I propose that modelling as an activity is only really justified by its relation to explanation. Because of this, the value of a model is tied to its structure as a tautology that corresponds to the known facts about the world, as outlined above.

Although there are diverse benefits of modelling, unless the driving purpose of a model is one of seeking to discover the implications of a theory (prediction) or to promote an explanation of some features of the world (sufficiency, insufficiency and existence proofs) then the modelling risks becoming a sterile exercise. Although models have a role in cultivating our intuitions, this must be taken beyond the level of the individual researcher and tested in the dissemination and contestation of the model by the scientific community.

Precisely because models are tautologies, the equations that comprise them don't have intrinsic meaning. It is not unambiguous which features of a set of equations that comprises a model are relevant to the theoretical issue at hand. Instead, the meaning of the model is created by the researchers who construct the model in their attempt to persuade others of their findings. When communicating their model and the implications which they derive from it, the obligation is upon the modeller to define which features of the model are supposed to correspond to which features of the world, and what the purposes of constructing a model like this are. Only when a proposal about the theoretical content of the model is offered can the value of the model be evaluated.

A good model is defined by the purposes you have; by whether you set out to deduce the consequences of the existence of certain entities to generate predictions, to test a proposal, prove necessity, sufficiency or insufficiency, or 'merely' to develop a formal framework for a problem. If the purposes of a model are not explicitly stated then its success and utility cannot easily be evaluated. Without a means by which the success of a model can be evaluated it will be difficult to integrate model findings in to a progressive programme of model development, something which is necessary for modelling to mature into a mature technique. (Roelofs, 2005)

## Conclusions

Starting with the accusation that models are merely tautologies, I have attempted to turn this accusation around and argue that models can be informative about the world, but only when the correspondences between their parts is carefully articulated. I have used the analogy of a tautology to suggest a framework for some of the benefits of modelling with relation to explanation. There remains a strict sense in which modelling does not provide 'new facts' about the world, but for this to remain a criticism relies on both an unrealistic view of the reliability of the facts of psychology and neuroscience as provided by other tools of the scientific method (Feyerabend, 1988) and on an overly conservative definition of what a 'new fact' is. If I use an equation to calculate the size of the earth from the length of a shadow at a particular spot at a particular time then I would say that I have discovered a new fact, even though this fact — the size of the earth — was inherent in the old facts of shadow length, time and position and nothing was added but the tautology of an equation.

This framework I have tried to outline of the benefits of modelling is not supposed to be comprehensive, but it does, I suggest, capture the core benefits of modelling, which are those relating to explanation. It is in these ways that modelling can help confirm, create, enhance or refute theories. Models aid explanation in the same way as mathematics: by enhancing our perception beyond the horizon of individual reason and intuition.

## References


Craik, K. J. W. (1943). *The nature of explanation*. Cambridge: Cambridge University Press.

Crick, F. (1989). The recent excitement about neural networks. *Nature*, *337*(6203), 129-132.

Dror, I. E., & Gallogly, D. P. (1999). Computational analyses in cognitive neuroscience: In defense of biological implausibility. *Psychonomic Bulletin & Review*, *6*(2), 173-182.

Ellis, R., & Humphreys, G. (1999). *Connectionist psychology : a text with readings*. Hove, UK: Psychology Press Ltd.

Elman, J. L. (1996). *Rethinking innateness: A connectionist perspective on development*. MIT Press.

Feyerabend, P. (1988). *Against method (revised edition)*. New York: Verso.

Hurlbert, A., & Poggio, T. (1988). Making machines (and artificial intelligence) see. *Daedalus*, *117*(1).

Krugman, P. (1993). How I work. In *the unofficial paul krugman archive.* www.pkarchive.org accessed 1/7/08.

Kukla, A. (1995). Amplification and simplification as modes of theoretical analysis in psychology. *New Ideas in Psychology*, *13*, 201-217.

Kulka, A. (2001). *Methods of theoretical psychology*. Cambridge, MA.: MIT Press.

Lewandowsky, S. (1993). The rewards and hazards of computer-simulations. *Psychological Science*, *4*(4), 236-243.

Linsker, R. (1988). Self-organization in a perceptual network. *Computer*, *21*, 105-117.

MacKay, D. J. C. (2003). *Information theory, inference and learning algorithms*. Cambridge, UK: Cambridge University Press.

Marr, D. (1982). *Vision: A computational investigation into the human representation and processing of visual information*. Henry Holt and Co., Inc. New York, NY, USA.

Mayes, G. (2008). Theories of explanation. In *The internet encyclopedia of philosophy.* www.iep.utm.edu accessed 1/7/08.

McClelland, J. L., & Rumelhart, D. E. (1981). An interactive activation model of context effects in letter perception: Part 1. an account of basic findings. *Psychological Review*, *88*, 375-407.

McCloskey, M. (1991). Networks and theories - the place of connectionism in cognitive science. *Psychological Science*, *2*(6), 387-395.

O'Reilly, R. C., & Munakata, Y. (2000). *Computational explorations in cognitive neuroscience: Understanding the mind by simulating the brain*. MIT Press.

Plaut, D. C., & Shallice, T. (1993). Deep dyslexia - a case-study of connectionist neuropsychology. *Cognitive Neuropsychology*, *10*(5), 377-500.

Popper, K. (1968). *The logic of scientific discovery*. London: Hutchinson.

Roberts, S., & Pashler, H. (2000). How persuasive is a good fit? a comment on theory testing. *Psychological Review*, *107*(2), 358-367.

Roelofs, A. (2005). From Popper to Lakatos: A case for cumulative computational modeling. In A.Cutler (Ed.), *Twenty-first century psycholinguistics: Four cornerstones* (Vol. 313-330). Mahwah, NJ: Lawrence Erlbaum.

Rumelhart, D., McClelland, J., & the PDP Research Group. (1986). *Parallel distributed processing: Explorations in the microstructure of cognition*. Cambridge, MA: The MIT Press.

Rumelhart, D. E., & McClelland, J. L. (1982). An interactive activation model of context effects in letter perception: Part 2. the contextual enhancement effect and some tests and extensions of the model. *Psychological Review*, *89*, 60-94.

Segalowitz, S., & Bernstein, D. (1997). Neural networks and neuroscience: What are connectionist simulations good for. In *The future of the cognitive revolution.* Oxford University Press.

Smolensky, P. (1988). On the proper treatment of connectionism. *Behavioral and Brain Sciences*, *11*(1), 1-23.