

How do we use computational models of cognitive processes?

T. Stafford

*Department of Psychology, University of Sheffield,
Sheffield, S10 2TP, UK*

** E-mail: t.stafford@shef.ac.uk*

<http://tomstafford.staff.shef.ac.uk/>

Previously I outlined a scheme for understanding the usefulness of computational models.¹ This scheme was accompanied by two specific proposals. Firstly, that although models have diverse purposes, the purposes of individual modelling efforts should be made explicit. Secondly, that the best use of modelling is in establishing the correspondence between model elements and empirical objects in the form of certain 'explanatory' relationships: prediction, testing, existence proofs and proofs of sufficiency and insufficiency. The current work concerns itself with empirical tests of these two claims. I survey highly cited modelling papers and from an analysis of this corpus conclude that although a diverse range of purposes are represented, neither being accompanied by an explicit statement of purpose nor being a model of my 'explanatory' type are necessary for a modelling paper to become highly cited. Neither are these factors associated with higher rates of citation. The results are situated within a philosophy of science and it is concluded that computational modelling in the cognitive sciences does not consist of a simple Popperian prediction-and-falsification dynamic. Although there may be common principles underlying model construction, they are not captured by this scheme and it is difficult to imagine how they could be captured by any simple formula.

Keywords: Computational Modelling; Cognitive Science; Philosophy of Science

1. Introduction

We might expect computational modellers to be very concerned with theory and meta-theory. For one reason, computational modelling is a relatively young branch of psychology and neuroscience. Not only this, but it is a field in which innovation abounds, as the rise and rise of computational power opens up new possibilities. Historically, this kind of tumult has been associated with discussion of the scope and purpose of a discipline, and with discussion of the standards of comparison that should be applied to different investigations. A second reason we might expect computational

modellers to concern themselves with theory and meta-theory, is that modelling generates no data in itself. Modellers are forced to exist in the world of theory; to simulate the underlying structures responsible for the patterns in the data, to propose different explanations for the data and to test relationships between proposed theoretical entities in our computational mini-worlds.

For these reasons we might expect computational modellers to resist the urge to view their work as a mere technical challenge, but remain alive to the theoretical claims that modelling work must be situated among if it to be scientifically meaningful. For the same reasons, we might expect computational modellers to be alive to the ongoing metatheoretical questions that concern computational modelling: what scientific role can modelling play, how should computational models be evaluated and what are legitimate motivations for instigating a computational modelling project? These kinds of questions are the domain of the philosophy of science.

Philosophy of science has a mixed reputation among scientists. It has been said that there is a remarkable disparity between the actual conduct of science and the picture presented by mainstream philosophy of science. One reason is the greater attention paid in philosophy of science to how science ought to be conducted — that is, to the logical requirements and structure of scientific claims — rather than how it is — in fact — conducted. The physicist Richard Feynmann is reported to have said “Philosophy of Science is as useful to scientists as ornithology is to birds”. The implication being that philosophy of science is an artificial and wholly conceptual domain of knowledge which is irrelevant to the way scientists conduct themselves. It would be surprising, however, if scientists were able to “do” science with quite the same instinct, grace and spontaneity that birds fly (even Feynmann).

There are at least three good reasons that philosophy of science is not just of interest, but a necessity for scientists. What is more, these three reasons are especially pertinent to the field of computational modelling. Firstly, an articulation of principles is required to support the acceptance of a new field. In the case of computational modelling it is not the case that everyone accepts its value as a scientific activity. For example, neuroscientist and Nobel Laureate Francis Crick accused neural network modelling of being ‘a rather low-brow enterprise’² and a vent for frustrated mathematicians. Segalowitz and Bernstein³ were clearer but equally condemnatory in their criticism, dismissing modelling and explaining that ‘models cannot tell us anything about the world...nor can they provide new information

about brain organisation or function'. Although these criticisms concern the historical period when computational modelling was still struggling for acceptance in psychology and neuroscience, it is still possible to find similar sentiments, along the lines that modelling is an indulgence or irrelevance, expressed informally today. Indeed, the division of psychologists or neuroscientists into 'modellers' and 'non-modellers' suggests that modelling has not been fully integrated into the wider discipline.

Secondly, philosophy of science informs debates that we have within a field. Among modellers substantial disagreements exist concerning the correct approach to modelling. As evidence of this assertion, let me pick two discussants from a debate that occurred on the comp-neuro mailing list in 2008. James Bower expressed the opinion that modellers should perhaps "give up on cerebral cortex for several hundred years and all study tritonia instead" .⁴ Although he was making this suggestion to illustrate a point, it does resonate with his apparent preference for low-level 'computational neuroscience' modelling. In contrast, in the same debate, Randall O'Reilly wrote that "the hippocampus is essentially a "solved problem" in terms of the general framework for how its biological properties enable its well-established role in memory" .⁵ Not only should we, *contra* Bower, continue to study cerebral cortex, but we have in fact essentially solved a major part of it! Bower initiated the discussion on the email list to illustrate to a group of graduate students that many fundamental issues within computational neuroscience are not agreed upon. This it illustrated admirably and the reader is encouraged to review the discussion to enjoy wide-ranging consideration of levels of modelling, ways of assessing the value of a model, the value of modelling in general. All of these issues are the business of philosophy of science and a modeller-scientist cannot avoid having a position of them, albeit if only implicitly.

Thirdly, and finally, a philosophy of science is necessary to educate the next generation of scientists. Even if we could do science as instinctively as birds fly, we would still wish to articulate the philosophy underlying our practice of science so that we could best convey it to future scientists. It is perhaps surprising, then, that of four major textbooks in computational neuroscience and psychology ⁶⁻⁹ very little space is devoted to the topic of what role computational models play in science. Perhaps the silence of the textbook authors is in recognition of the seemingly-intractable nature of many debates in philosophy of science, and a consequent desire to avoid unfruitful discussion. I recognise the risk that claims concerning the philosophy of science may evoke counter-claims and so on *ad infinitum*. In the

current paper, my discussion of the purposes of computational modelling is grounded by a survey of how computational models are presented in the literature. In this way I hope to combine consideration of how computational modelling *should* proceed with consideration of how it *does*, in fact, proceed.

2. Three claims about computational models

Previously, I have proposed a scheme for categorising the purposes of computational models.¹ The details of this scheme are less important, for the purposes of this paper, than three claims which I will use here to motivate the current work. The first claim is that there are many purposes for which you might build a computational model. This is a reflection of the fact that it is difficult to elaborate a single formula which captures what all modellers are trying to achieve with every model. Therefore, it is likely the case that different modellers are trying to achieve different things, and so there must be many purposes for which computational models are built. The second claim, which recognises the first, is that models ought to be accompanied by some statement of what the modeller hopes to achieve by that model. If models can have many different purposes, then appropriate assessment of a model will take account of those purposes for which a model is designed. And this is made easier if the model-builder reveals their purposes rather than leaving them to be inferred. The third claim of Stafford (2009) is that, although there are many purposes for model building, the best purposes are those which relate to providing explanations. This is the claim that models that use correspondences between model parts and real-world entities to make, refine or test predictions. In this claim I am influenced by Popperian philosophy of science¹⁰⁻¹² and the auxiliary assumption that modelling is a kind of theory construction. This assumption makes natural the application of the centrality of prediction and falsification from Popperian philosophy of science to computational modelling.

The scheme for categorising model purposes developed by myself,¹ and extended here, is shown in Table 1. There are four major categories of model purposes according to this scheme. Exploratory model building includes the sub-categories *capacity*, which is the demonstration that a model has the capacity to perform a certain kind of function, without reference to how that model might relate to psychological or neuroscientific theory. For example, the demonstration that a Hopfield network can store patterns would be such a demonstration of capacity. *Data fitting* is the demonstration that a model can generate data which resembles the data generated in psychology

or neuroscience investigation. *Biological plausibility* is the adjustment of an existing model to increase the extent of its correspondence to the biological structure it purports to model. *Reinterpretation* is the use of suggestive results from a model to widen the scope of plausible explanations. *Problem definition* is the use of modelling to explore and more fully define the domain in which a psychological or neuroscientific function is performed.

Table 1. Model paper categorisation scheme.

Major category	Sub-categories
Exploratory	Capacity Data fitting Biological plausibility Reinterpretation Problem-definition
Analysis	—
Integrative	—
Explanatory	Prediction Testing Sufficiency Existence proof Insufficiency

The second and third major categories are models used for *analysis* (e.g. a statistical model such as a linear regression) and modelling for *integration*, which is the construction of a model which combines models from two separate domains or levels of description.

The fourth and final category, which I claim is the one that the most scientifically useful models belong to, is of models with *explanatory* purposes. To understand my breakdown of this category I will need to rehearse an argument made previously,¹ which attempts to understand explanation in terms of the “modelling is just tautology” accusation quoted earlier.³ My argument, briefly, was that models must, in some sense, be only tautology but they derive their power from the correspondence between the parts of the model and real-world entities. All mathematical equations are tautological, but this does not mean that computation cannot be used to reveal new facts about the world. If you take the length of the shadow of a tower at noon in one place, and the length of the shadow of a tower at another place at noon you can compute the circumference of the earth. The result is an inherent and necessary result of the information you put into

the computation. In this sense it is tautological, but it would be obtuse to argue that the computation has not revealed new information about the world.

We can take the simple example of $1 + 2 = 3$ — another tautology — and use it to illustrate the value of modelling-as-tautology. If the model elements on the left-hand side of the equation ('1 + 2') correspond to known real-world entities then the model predicts the presence of the entities that correspond to the right-hand side elements ('3'). If entities corresponding to both left- and right-hand side elements are known that the model demonstrates that the left-hand side elements are *sufficient* to produce those entities corresponding to elements on the right-hand side. If, alternatively, the entities known to exist are more than represented by elements on the right-hand side (for example, not '3' but '4' maybe) then the model constitutes a demonstration of *insufficiency* (particularly, of those entities represented by the left-hand side of the equation to produce those entities on the right). If the existence of the entities corresponding to elements of the model is in doubt, or the particular interrelation represented by the model is in doubt, then the model can constitute a form of *existence proof* that these entities can exist in the particular inter-relation captured by the model. These four types of explanation, which correspond — I suggest — to the canonical Popperian category of prediction and three examples of what Kukla^{13,14} calls *theory amplification* make up the four subcategories of my fourth class of model purpose. Note that the categorisation of a model depends on its relationship to wider theory, not on its internal structure.

3. The current work

3.1. Aims

The current work is concerned with an empirical investigation of how computational modelling work is presented to the scientific community. My hope was that a set of highly-cited modelling papers would act as a proxy for successful or admirable modelling work. By systematic investigation of the properties common to this set we might get some insight into the characteristics of modelling papers that are associated with success (as defined in terms of high rates of citation). The scheme outlined above is used to categorise the modelling papers investigated, so this investigation also acts as a test of the adequacy of this scheme for categorising modelling paper type.

3.2. Corpus

Table 2. Search terms used to identify modelling papers from each source

Source	Search criteria
Nature	Topic=(computational) AND Topic=(neuroscience OR psychology)
Nature Neuroscience	Topic=(computational)
Neural Computation	ALL
Cognitive Science	Topic=(model)
Connection Science	ALL
NCPW11	ALL

The papers selected for this survey were the fifty most highly cited modelling papers from five journals, plus the papers from the 11th *Neural Computation and Psychology Workshop* (NCPW) held in 2008.¹⁵ The journals were selected to contain a range of papers for a specialist and generalist audience, and to capture some difference in impact factor. The search terms used to identify modelling papers are shown in Table 2. Non-modelling papers identified by these searches were discarded without replacement. Papers with less than 10 citations were also discarded, resulting in a total number of papers included in the survey of 173.

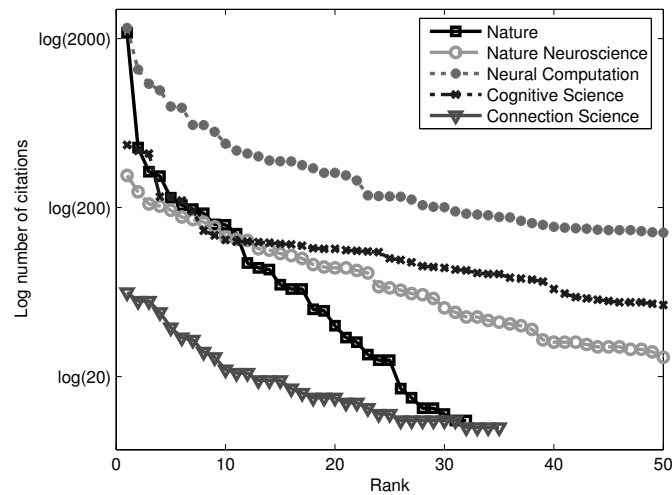


Fig. 1. Log citations for the fifty highly cited papers included in the survey

Figure 1 shows the log of the number of citations of the papers initially selected for inclusion in the survey, from the five journal sources, using the search times given in Figure 2. As expected, journals with higher impact factors have more highly cited papers.

4. Results

4.1. *Is making explicit your purpose for building a model associated with publication in quality journals and/or higher citation counts?*

To address this question each paper in the corpus was coded as to whether it made explicit in the abstract what the purpose of the modelling work presented was. The assumption here is that, because models have diverse purposes, if it is not said why a model is built then that model cannot easily be assessed or used by the non-modelling community.

For the NCPW11 conference, the proportion of papers which, in their abstracts, were explicit about the purpose for which their model was constructed was 66%. In other words, most, but far from all, models were explicit about their purpose. This finding confirms an informal observation that I made while at NCPW11. One researcher I spoke to during NCPW11 acknowledged that this state of affairs was sub-optimal, but expressed the opinion that it was due to the nature of the papers at a conference. In other words, this was provisional work. Papers accepted for publication in a journal would have a far higher proportion of those which were explicit about their purposes. The results of my survey, shown for all the sources in the corpus, are shown in figure 2.

The details of these results are discussed below. Because it seemed that in general that most, but by no means all, papers were explicit about their modelling purpose, an additional analysis was carried out. The mean of the citation counts was calculated for each source, divided according to those papers which were explicit about their purpose and those that were not. If, on average, papers which were explicit were more highly cited (even among this corpus of highly cited papers) then this analysis should reveal it. The results are shown in figure 3. There is no clear superiority, in terms of citations counts, of the 'explicit' papers compared with the 'non-explicit'.

4.2. *What purposes are associated with highly cited papers?*

The full results of the survey with respect to the primary purpose of the modelling paper analysed, are shown in table 3. Each paper can contribute

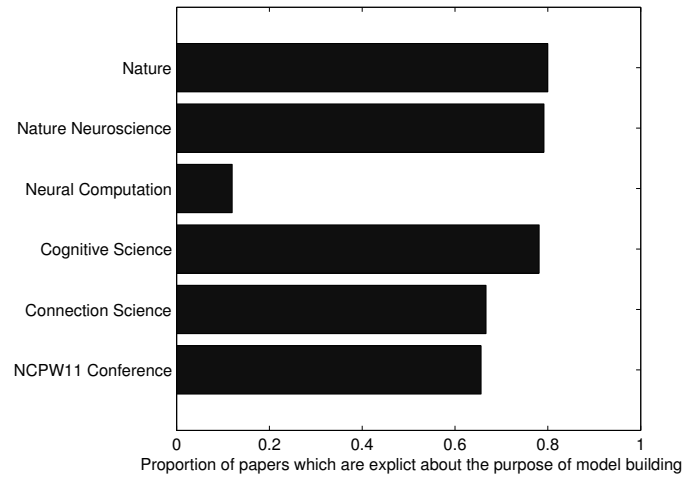


Fig. 2. Proportion of papers with an explicit statement of purpose in the abstract, by source

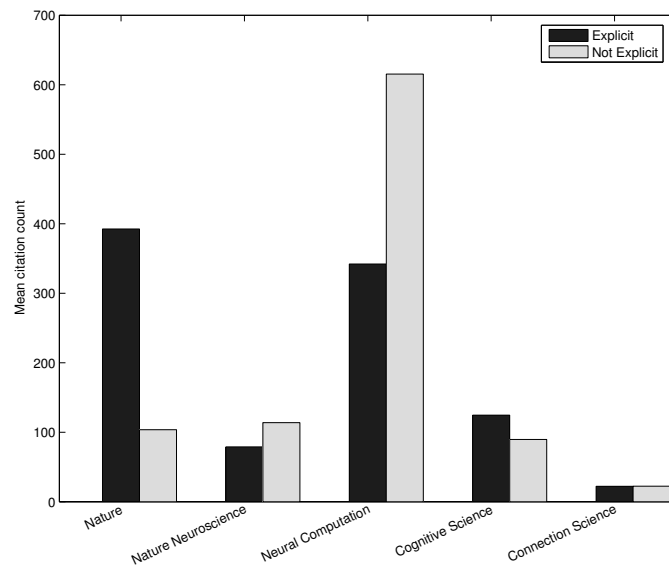


Fig. 3. Mean citation count for papers with and without an explicit statement of purpose in the abstract, by source

10

to only one cell.

Table 3. Survey results, model paper types by source.

Major category	Minor category	Nature	Nature Neuroscience	Cognitive Science	Connection Science	Neural Computation	NCPW11	TOTAL
Exploratory	Capacity	2	5	4	18	22	17	68
	Data fitting	0	1	6	0	0	6	13
	Biological plausibility	0	0	1	2	0	5	8
	Reinterpretation	1	0	1	0	0	2	4
	Problem-definition	0	0	1	1	0	0	2
Analysis Integrative Explanatory	—	2	1	2	0	0	2	7
	—	1	5	5	7	3	6	27
	Prediction	1	4	3	1	1	3	13
	Testing	0	0	6	2	0	1	9
	Sufficiency	2	7	3	3	0	2	17
	Existence proof	0	2	0	0	0	0	2
	Insufficiency	1	0	0	2	0	0	3

Note that there are over twice as many ‘exploratory’ than ‘explanatory’ papers. Of the exploratory papers the majority are of the ‘capacity’ kind. The second largest subcategory of the exploratory type is ‘data fitting’. These two categories are very similar in nature, but distinguished in that ‘capacity’ papers demonstrate that a model can do some abstract or general task, whereas ‘data-fitting’ papers demonstrate that a model produces data of the same form as some experimental procedure.

Note also that the full range of proposed modelling purposes is found by the survey. In other words, there are no empty cells in the categorisation table (although there are some empty cells with respect to individual sources in the corpus).

In order to further address the question of the superiority of explanatory modelling, compared with models built for other purposes, the proportion of models with explanatory purposes for each source in the corpus was calculated. The results are shown in figure 4. It is clear that a majority of papers, in nearly all sources, are **not** presented as fulfilling explanatory purposes. Even for the single source for which more than half of papers in the corpus were explanatory, the proportion was not very much greater than half.

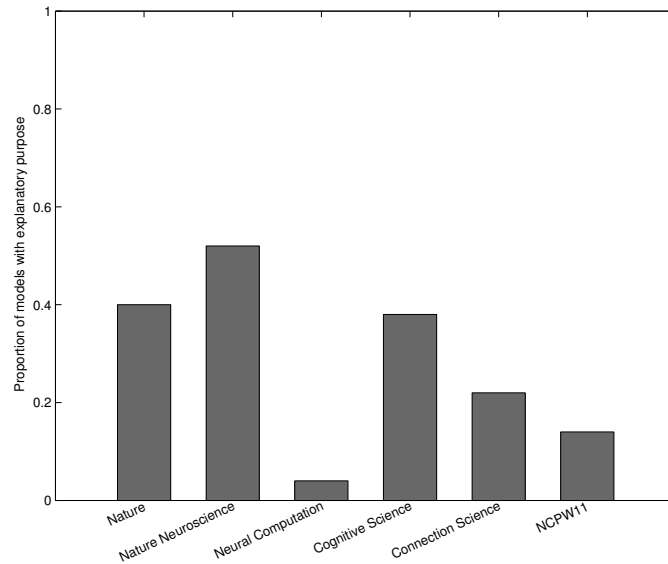


Fig. 4. Proportion of models with an ‘explanatory’ purpose, by source

An analysis of mean citation counts for explanatory compared with non-explanatory papers from each source is shown in figure 5.

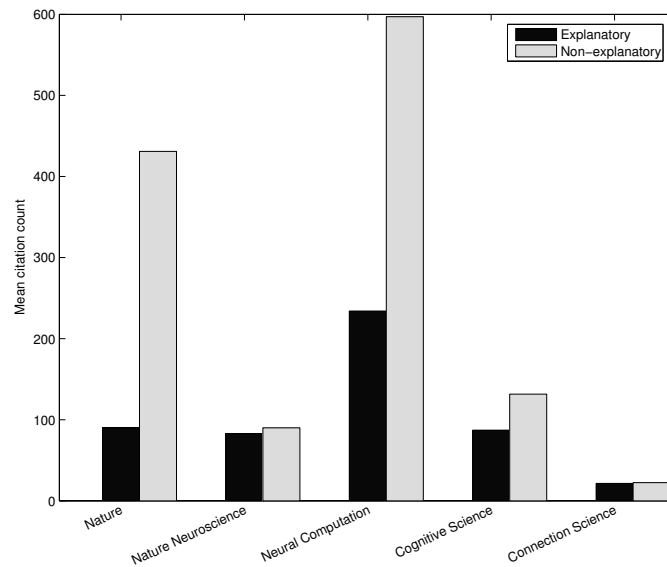


Fig. 5. Mean citation count for ‘explanatory’ and ‘non-explanatory’ modelling papers, by source

5. Discussion

The results covered in the previous section put us in a position to address the claims asserted previously and discussed at the beginning of this current paper (section 2). The first claim is that there are many purposes for which computational models are built. The results confirm this. Across the entire corpus examples of each purpose in the categorisation scheme were found. Two categories were well-populated, against previous expectations. The importance of modelling for providing theory integration and novel frameworks (as reflected in the ‘integration’ category), and the importance of modelling for providing new methods/techniques (as reflected in the ‘capacity’ category) was unanticipated from the perspective of my previous analysis.¹

The second claim is that models ought to be accompanied by an explicit statement of what the modeller hopes to achieve by that model. These data

cannot address this claim, since it is normative by nature, but they can inform us as to what occurs ‘in the wild’ with respect to model publishing. Evidently, many highly cited papers are not accompanied, in their abstracts, by an explicit statement of the purpose for which they are built. Further, it does not seem as if models accompanied by an explicit statement of purpose have a higher citation count among the corpus, on average (Figure 5). Although the mean citation count is higher for ‘explicit’ models for those published in *Nature* (see figure 3), the opposite pattern was true for models published in *Neural Computation*. Furthermore, the highly skewed distribution of citation counts (see figure 1) means that a small number of highly cited papers have a disproportionate impact on these mean figures, and so although the difference between the ‘explicit’ and ‘non-explicit’ means seems large, it is probably not reliable. This inference is supported by the fact that although the differences are large for both *Nature* and *Neural Computation* journals, they are in opposite directions. Although I would still support the normative claim that models *should* be accompanied by an explicit statement of their purpose, this support is not reinforced by the data presented here. Many successful modelling papers do not contain such a statement.

The third claim is that the best model purposes are those which relate to providing explanations. Considering the proportion of explanatory papers from each source, it does not seem as if there is a strong association between better quality journals and a higher proportion of explanatory papers. Three higher impact journals, *Nature*, *Nature Neuroscience* and *Cognitive Science* do seem to have elevated levels of explanatory papers, compared with the NCPW11 conference and *Connection Science*. An exception to this pattern is the *Neural Computation* journal, which has a very high impact factor, but a very low proportion of papers with explanatory purpose. The reason for this can be deduced from table 3. Nearly all the papers in *Neural Computation* are demonstrations of capacity.

Note, however, that for all sources the proportion of explanatory papers is low. The mean citation counts for explanatory compared with non-explanatory papers shows there is no evidence that explanatory papers have higher rates of citation. If anything, there is some evidence that non-explanatory papers have higher rates of citation, although the distribution of citations (as discussed above) could make the means unrepresentative, and the results for the *Neural Computation* journal should probably be excepted. (The differences between the type of papers published in *Neural Computation* and the other sources in the corpus are probably due to the

fact that it is a specialist journal with an engineering slant, rather than a general scientific or cognitive science journal like the others in the corpus).

A large minority of papers fell into the ‘integrative’ category, something which was unanticipated from my initial theorising, although it is in line with Kukla’s analysis of the scope of theoretical psychology (alongside which I would include modelling) .¹⁴

It is surprising, perhaps, that so many of this corpus of highly cited modelling papers are of the ‘data-fitting’ category. An influential review by Roberts and Pashler ¹⁶ condemns data-fitting as a criterion for model assessment.

5.1. *Final words*

A limitation of the current design is that the categorisation of papers was done by one person (myself). The use of a single reviewer means it is impossible to assess the reliability of the categorisation. A possible extension of this work would be to fully formalise the criteria for the categories used and to have papers categorised and rated with respect to whether they include an explicit statement of purpose by independent reviewers who were blind to the source and authors of the paper.

Nonetheless, even allowing for some minor to moderate level of intra- and inter-reviewer variability, the major conclusions of this review would hold true: highly cited modelling papers appear to be constructed for a wide variety of theoretical purposes, they are often not explicit about what their purposes are, and often these purposes are not ‘explanatory’ according to this scheme. Neither being explicit nor having an explanatory purpose appear associated with higher rates of citation. This suggests that although previously I have suggested that these are desirable properties of a modelling paper, their absence does not conspicuously hinder the reception of a modelling paper.

The results presented here could be further investigated by looking at modelling papers cited outside of modelling journals — in other words, by experimentalists. This would give a valuable insight into how modelling work affects mainstream cognitive science. Another productive avenue would be to look how modelling papers build on and test existing models. The cumulative nature of research programmes, and how theories succeed or are replaced, is another area where analysis of the nature of computational modelling could be informed by philosophy of science .^{17–19}

The current review is an investigation of how science, at least in this corner of the domain, *is* carried out, rather than how it *should be* carried

out. In this sense the review is in the spirit of Feyerabend²⁰ and seems to echo his conclusion that “anything goes”. He used this phrase to summarise his conclusion that there are no principles which hold universally in the conduct of scientific investigations. Here we might take it in a weaker sense to reflect the conclusion that successful modelling papers are not of one type and their nature is not captured by my initial hypothesis about what makes a good modelling paper (i.e. explicit statement of purpose, explanatory purpose).

A reasonable extension of this conclusion would be that ‘naïve Popperianism’¹¹ is demonstrably wrong, at least in this domain of science. There is far more to computational modelling in the cognitive sciences than prediction and falsification. Although we may still hope that there are general principles governing the desirable features of modelling papers, this review suggests that they are not captured by this level of analysis. General principles, if it is possible to find and articulate them, are likely to be complex. Like all modellers, I continue in the belief that modelling is a powerful scientific tool. But if modelling is a tool it is clear that it is a multipurpose tool, used by different scientists in different ways.

6. Acknowledgments

Thanks The Redwood Center for Theoretical Neuroscience at UC Berkeley for hosting me while I did this work, and Bruno Olshausen and Nicol Harper for facilitating my visit there, to Eddy Davelaar for advice before and during preparation of this paper, and to an anonymous reviewer for a thorough reading of a draft of this paper.

References

1. T. Stafford, What use are computational models of cognitive processes?, in *Connectionist Models of Behavior and Cognition II: Proceedings of the 11th Neural Computation and Psychology Workshop*, ed. N. R. . K. P. J. Mayor (World Scientific, 2009) pp. 265–274.
2. F. Crick, *Nature* **337**, 129 (1989).
3. S. Segalowitz and D. Bernstein, Neural networks and neuroscience: What are connectionist simulations good for, in *The future of the cognitive revolution.*, (Oxford University Press, 1997)
4. J. M. Bower. Email to comp-neuro@neuroinf.org, 12/08/2008, 16:47 GMT available from <http://www.neuroinf.org/pipermail/comp-neuro/2008-August/thread.html>.
5. R. C. O’Reilly. Email to comp-neuro@neuroinf.org, 12/08/2008, 18:19

GMT available from <http://www.neuroinf.org/pipermail/comp-neuro/2008-August/thread.html>.

6. R. C. O'Reilly and Y. Munakata, *Computational Explorations in Cognitive Neuroscience: Understanding the Mind by Simulating the Brain* (MIT Press, Cambridge, MA., 2000).
7. P. Dayan and L. Abbott, *Theoretical neuroscience: Computational and mathematical modeling of neural systems* (MIT press Cambridge, MA., 2001).
8. R. Ellis and G. Humphreys, *Connectionist Psychology : a text with readings* (Psychology Press Ltd, Hove, UK, 1999).
9. J. L. Elman, *Rethinking Innateness: A Connectionist Perspective on Development* (MIT Press, 1996).
10. K. Popper, *The Logic of Scientific Discovery* (Hutchinson, London, 1968).
11. B. Magee, *Popper* (RoutledgeFalmer, 1974).
12. A. Chalmers, *What is this thing called science?* (Univ. of Queensland Press, 2006).
13. A. Kukla, *New Ideas in Psychology* **13**, 201 (1995).
14. A. Kukla, *Methods of Theoretical Psychology* (MIT Press, Cambridge, MA., 2001).
15. J. Mayor, N. Ruh and K. Plunkett (eds.), *Connectionist Models of Behavior and Cognition II: Proceedings of the 11th Neural Computation and Psychology Workshop* (World Scientific, 2009).
16. S. Roberts and H. Pashler, *Psychological Review* **107**, 358 (2000).
17. A. Roelofs, From Popper to Lakatos: A case for cumulative computational modeling, in *Twenty-first century psycholinguistics: Four cornerstones*, ed. A. Cutler (Lawrence Erlbaum, Mahwah, NJ, 2005)
18. I. Lakatos, Falsification and the methodology of scientific research programmes, in *Criticism and the Growth of Knowledge*, eds. I. Lakatos and A. Musgrave (Cambridge University Press, Cambridge, 1970) pp. 91–195.
19. T. Kuhn, *The structure of scientific revolutions* (University of Chicago Press, Chicago, 1996).
20. P. Feyerabend, *Against Method (revised edition)* (Verso, New York, 1988).