Title
Tools and Techniques for Quantitative and Predictive Cognitive Science

Permalink
https://escholarship.org/uc/item/1860d7qb

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 28(28)

ISSN
1069-7977

Author
Stewart, Terrence C.

Publication Date
2006

Peer reviewed

Tools and Techniques for Quantitative and Predictive Cognitive Science

Terrence C. Stewart (terry@ccmlab.ca)
Carleton Cognitive Modelling Lab, Institute of Cognitive Science, Carleton University
1125 Colonel By Drive, Ottawa, Ontario, K1S 5B6, Canada

Abstract

A methodology is described for developing cognitive science theories which produce numerical predictions. This is done by adopting methodology from mathematical models in physics, and adapting it for use with the more complex computational models. Bootstrap confidence intervals and equivalence testing are introduced, and parameter fitting is shown to be an intermediate step before prediction. To ensure replication and exploration by other researchers, publication of the source code for the model, experimental situation, and data analysis is required. To assist in this process, we have developed a freely available tool suite, covering creation of models, running parallel simulations, parameter exploration, data analysis, and Internet-based access to all data.

Introduction

As cognitive science theories become more complex, it is increasingly difficult for the predictions of those theories to be determined. The cognitive behaviour being examined is a result of an exceedingly complicated interconnection of components, each of which is itself complex. This makes the interpretation of the results of cognitive science research somewhat unclear. Are we learning about fundamental mechanisms of cognition, or are we describing their effects? Are we finding causal links, or correlations? Are we developing frameworks, or playing "twenty questions" with nature? And what, exactly, is the difference between these possibilities?

One methodology that is commonly used to help frame this research is *computational modelling*. Here, a key component of the research is the creation of a computer program which is a *model* of the actual cognitive system under investigation. However, following this approach does not, in itself, resolve the above questions. Is the computer program meant to be an exact expression of the theory? Is it meant to indicate actual processes occurring inside the system? Is it predictive or descriptive? How do we interpret the parameters of the model? Are we modelling individuals or the group mean? How can we evaluate how well our models perform?

These sorts of questions must be addressed in any research. Since computational modelling is a relatively new approach for science, it is especially important to be clear on what claims are being made, what aspects are being measured, and what, exactly, constitutes proof. The goal of this paper is to present a philosophical framework for understanding how computational modelling fits within science, and then from that basis to describe tools and techniques that help provide rigor to this scientific approach. These principles and processes have been fundamental to our work in the Carleton Cognitive Modelling Lab, and we believe they have wide applicability and utility elsewhere.

Standard Practices

Examining the published results of computational modelling in cognitive science reveals a wide variety of approaches and measurement standards. For example, the following three studies demonstrate various approaches for identifying correspondences between a complex computational theory and the observations that theory is meant to explain..

(1) In (Erev & Barron, 2005), a model of forced-choice decisions is studied by measuring the mean squared deviation between the choice rates for the model and human subjects. Parameter fitting is used to adjust the model to find the closest match to human data. After this, a further comparison is done to a separate set of human data, so as to evaluate the model's ability to generalize.

(2) In (Connell & Keane, 2006), a model of plausibility judgments is investigated by measuring the correlation between the model's output for 60 different situations and the mean human judgment for the same situations. Sensitivity analysis is used to show how this correlation changes for differing parameter values of the model.

(3) In (Anderson et al., 2005), an ACT-R model of the Tower of Hanoi task is examined. The model is first fit based on human latency information, and is then used to generate probability distributions for the BOLD responses in three brain regions. These predictions are evaluated via a χ^2 goodness-of-fit to human fMRI results, and a non-significant result leads to the conclusion that "the model does account for the systematic variance in the data".

There are certainly many other approaches, and a detailed analysis of the benefits and drawbacks of each would be a monumental task. However, no single technique will completely satisfy everyone. Researchers try to choose the technique which best suits their needs, but it is difficult to foresee all questions a reader may have.

This problem becomes even more apparent in conference proceedings, where space limitations lead to a drastic reduction in provided information. Here, it is common to provide a simple side-by-side comparison of the model data and the mean of the subject data, after parameter fitting on a measure such as the mean-squared error. Distributional information and confidence intervals are seldom provided.

It is important to note that these observations and the ideas in this paper should *not* be interpreted as a criticism of individual papers in this area. Instead, this article is intended as a re-evaluation of customary research and publishing practices. It is difficult to produce concise and convincing measures for these varying situations, and space constraints (whether in conferences or journals) make an exhaustive approach impossible. To address this, we now discuss a number of limitations of common modelling practices, followed by our alternative methodology addressing these limitations. We then describe our software tools which simplify the process of following this approach.

Measurement Limitations

The most straight-forward limitation of these standard practices involves measurement. The general lack of confidence interval information makes it difficult to assess the actual degree of match. While these intervals are not always strictly necessary for the model data (as models can be run large numbers of times, giving a large N, and thus a small confidence interval), the real-world data used for comparison almost always has a much lower N. This means that the exact measured mean value *is not a meaningful number*. A close fit to this number is merely a close fit to the peculiarities of that particular sample of the overall population. Rather, we should be more interested in the confidence interval, which indicates that we are fairly sure (usually 95%) that the actual mean value (if we sampled the entire population) is within some range.[1]

We must also be careful using confidence intervals for comparisons. It is true that if two confidence intervals do not overlap, then there is a statistically significant difference between the sets of data. This is can be correctly used to conclude that the model does not match with the real world data[2]. However, if the confidence intervals *do* overlap (i.e. when there is no statistical significance), then we should make no conclusions at all. After all, if our criteria is that a good model is merely one which gives a confidence interval overlapping with the real data's confidence interval, then we are implicitly encouraging researchers to use a small sample size, leading to larger confidence intervals. Indeed this sort of analysis (or the equivalent t-test) should only be used for identifying *differences*, not *similarities.* For further discussion, see (Beaulieu-Prévost, 2006).

Furthermore, it is exceedingly rare to see any statistical measurement other than the mean being used. Since most cognitive behaviour has a high degree of variation, a good model of that behaviour *would also exhibit that same degree of variation*. Indeed, the real-world data and the model data should have indistinguishable distributions, not just indistinguishable means.

Fitting Limitations

A more important limitation concerns the process of parameter fitting, and is best summarized by Roberts and Pashler (2000). They highlight the fact that demonstrating that a model can be adjusted to fit a particular set of data *does not in itself inform you about the validity of the theory*. In particular, it says nothing about what range of real-world data the model could have been adjusted to fit. Perhaps by adjusting parameter settings it would possible to match the model to *any* plausible set of data. If this is the case, then demonstrating a good fit merely indicates that the model is highly adjustable, not that it captures some important aspect of the particular situation being modelled.

Another aspect which must be considered is that we should also have a way of knowing that when a particular model does fit well, that a variety of other models *do not fit as well*. For example, in (Stewart, West, and Coplan, 2004), we show that, for a certain set of measurements, a complex model of peer group interaction and friendship formation fits the real-world data no better than a completely random model. This demonstrates that studying a single model in isolation can lead to a false sense of the model's accuracy.

Communication Limitations

The final limitation to be considered involves the dissemination of information about the model.

> Sharing work has been so difficult that researchers tend to build their own animat minds and worlds from scratch, often duplicating work that has been done elsewhere.... Often, the only person who ever does experiments with an animat or agent is its author. In this field it has become acceptable not to have direct access to many of the major systems under discussion. (Humphrys & O'Leary, 2002)

Not only does this violate the basic tenant of replication in science, but it also leads to a situation where there are more types of models being investigated than there are current comparisons between architectures (Guilot and Meyer, 2000). Furthermore, when replication is attempted, we consistently find that vital aspects of the models *are not recorded in the paper describing them*. Axelrod (2005) describes his experience in a project attempting to replicate eight standard computational models in the social sciences as one where “Murphy's law seemed to be operating at full strength.” Aside from standard debugging issues, he found that there were ambiguities, gaps, and errors in the descriptions of models. Even when complete source code was available (a rare event), there were still problems involving the readability of the code and even such issues as the floating point accuracy of the computers the programs were being run on. All of these issues combine in such a way that it is difficult to evaluate published results of such research, and difficult to work with or expand upon the models developed by others.

This limitation also extends to the communication of the *results* of modelling work, as mentioned earlier, due to the constrained space available in publications.

Model-Based Science

To resolve these limitations and to develop a more rigorous approach to quantitative modelling in cognitive science, we need to take a closer look at the scientific methodology being applied. Some (Axelrod, 2005) have argued that using computation models and simulation is an entirely new way of doing science. Instead, our approach is to examine computational models as a generalization of standard mathematical modelling, as exemplified in physics. However, due to the increased complexity of the computational models, many of the simplifying characteristics that we have come to expect in physics models will not apply, forcing us to find alternate ways of dealing with old problems.

[1] Or, more correctly, if the actual value was outside that range, then we would only measure values as strange (or stranger) as what we did measure less than 5% of the time.

[2] Or, more correctly, if the model did exactly match the real data, then we would observe the sort of data we did observe less than 5% of the time.

Describing this approach requires us to be clear about exactly what we are trying to do as scientists. For our research, we do not believe that science is best described as the pursuit of truth. Instead, we are adopting Giere's argument in *Science Without Laws*:

> Rather than thinking of science as producing sets of statements that are true or false in the standard objectivist fashion, we should think of it as a practice that produces models of the world that may fit the world more or less well in something like the way maps fit the world more or less well. (Giere, 1999)

In other words, the goal of science is to develop set of rules (or *principles* or *theories*) which allow us to take a particular real-world situation, analyze it by measuring certain aspects, create a model from the results of that analysis, and then use that model to produce accurate predictions as to other aspects of that situation (such as its behaviour into the future). For the purposes of this paper we will not detail Giere's conclusion that this approach to the philosophy of science results in all of the features we want science to have (see Giere, 1988 and Giere, 1999), such as producing explanations as well as predictions.

This *model-based* science is an alternate way of describing the standard scientific approach, and one which leads to direct methodological solutions to the aforementioned problems common to cognitive modelling research.

Measurement Techniques

The key question for determining the appropriateness of a model in physics or in cognitive science is whether its predictions match those of the real situation being modelled. This match is performed by measuring some aspect of the real world, and measuring some aspect of the model, and comparing the two. The usefulness of the model is measured by how *closely* its predictions match the observed situation. In physics, much of the time these measures are highly non-variant: repeated measurements yield results similar to many decimal places.

However, in cognitive science, we do not have the luxury of only studying phenomena of low variability. Instead, our repeated measures of either the real data or the model data may look more like those in Table 1.

Table 1: A set of measurements

Sample Data			
1	2	2	3
2	1	0	5
2	3	3	2
1	1	1	3

Given this set of real, measured data, we want our models to produce *statistically equivalent* data. In other words, a good model should produce data with the same statistical distribution as we find in the real world. It should be noted that this is exactly what mathematical models in quantum physics do. To examine the distribution, we can make a number of different statistical measures (Table 2).

Table 2: Statistics for the sample in Table 1

Sample Mean:	2.0
Sample Median:	2
Sample Standard Deviation:	1.1726
Sample Skew:	0.69775
Sample Kurtosis:	3.5041

However, these statistics are measures of our *sample*, not the actual distribution. If we had the ability to have thousands or millions of individual measurements in our sample, then the sample distribution would approach the desired value. This is rarely the case in cognitive science research, meaning that if we build models that to match this particular sample's distribution, then we run the risk of over-fitting to that particular situation, and thus producing models which do not generalize.

The usual method for addressing this involves basing our comparisons on confidence intervals. However, standard approaches to confidence interval estimation are *parametric*: they make certain assumptions about the overall distribution of the data (such as it being Gaussian). Instead of making this assumption, we use the *bootstrap* method (Davison and Hinkley, 1997), which is known to be non-parametric, and thus leads to more accurate confidence intervals for a non-normally distributed data. It should be noted that this technique allows for a confidence interval for any measure, including such computationally intractable ones as the median.

Table 3: 95% Bootstrap Confidence Intervals

Mean:	1.4375 – 2.5625
Median:	1 – 3
Standard Deviation:	0.696 – 1.541
Skew:	-0.612 – 1.534
Kurtosis:	1.472 – 5.236

The confidence intervals give us a more accurate description of what is know about the real situation we are trying to model. We can also use them on the data produced by the model, saving us from having to run the model thousands of times before we can trust its statistics are representative.

However, we must be careful in applying these confidence intervals in this situation. As discussed previously, if we cannot say that if the confidence intervals of the real-world and model data overlap, then the model is good. Instead, we will make use of the relatively unknown statistical tool called *equivalence testing*.

Equivalence testing is a technique used in the evaluation of drug treatments to determine if a new, cheaper drug is as effective as some other drug, to within some pre-defined range. This is a modified version of the standard t-test, where instead of the traditional null hypothesis that the means of two groups are equal (μ_r-μ_m=0), the null hypothesis is that the difference between the means is greater than some amount ($|\mu_r$-$\mu_m|>\theta$). The value of θ defines the range of acceptable results. If we perform this statistical test, using μ_r as the real data set, and μ_m as the data from a given model, then a p-value less than 0.05 allows us to conclude with 95% certainty that the model and the real system do not differ by more than our threshold, θ. This approach can also be applied to ensuring that other statistical measures are also statistically indistinguishable.

The above description is intended for situations where we have some pre-determined threshold in mind, and we are looking for models that are at least that close. This generally not something that is available when first developing a model for a situation. In these cases, instead of setting the threshold and determining the p-value, we can instead set the p-value and determine the required threshold. This gives us a statistical measurement which has an intuitive interpretation. If we get a value of 0.1, then we are 95% certain that this model produces data that differs from the real data by no more than 0.1.

To demonstrate this alternate approach, consider the data shown in Figure 1. Here we have two sets of data and 95% confidence intervals for each. Under the assumption that the actual value for each measurement is within the confidence interval, the *maximum* difference between the two would occur when the left-hand measurement is at the top of its range, and the right-hand measurement is at the bottom. The difference between these values is the threshold for which the equivalence test would give a $p<0.05$ significance level. In other words, we can say that the model and the data are *statistically significantly similar* to within that range.

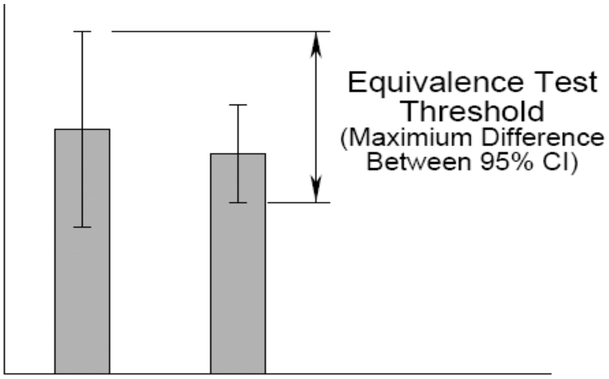

Figure 1: Real and model data with 95% confidence intervals. The equivalence test threshold is the maximum difference that could occur between the model data and the real data, assuming they are both within their respective 95% confidence intervals.

This gives us a measurement of the correspondence between the model and the real world. Unlike measures such as R^2, this is a directly interpretable measure. It tells us how close we can expect this model's predictions to be. Most importantly, it takes into account sampling error. It is this measurement that we believe should be the standard for modelling in cognitive science.

It is generally the case that we wish to make more than one prediction from a model. Usually, we have multiple data points from the real world, and we wish to see how well our model matches for all of these points, not just one. We also generally want to have predictions in multiple domains, such as recent work extending ACT-R to predict blood oxygenation levels as well as reaction times and error rates (Anderson et al., 2005). This broad applicability is a key criteria in physics for accepting a model as an explanation, not merely a predictive tool.

To attain a single value which expresses the model error across multiple predictions, we should avoid tools like the root mean squared error. This gives us a measurement indicating how much each measurement, *on average*, deviates from the mean. However, this can (and often does) obscure situations where one measurement significantly differs. If we are looking for predictive models across that set of measurements, then we should be more interested in the *worst* the model does at predicting, not its *average* fit across the particular measures chosen. For these reason, we advocate combining equivalence test threshold measures by taking the maximum error, rather than the average error.

Fitting Techniques

Mathematical modelling also provides us with a mechanism for addressing the parameter fitting problem. The models developed for cognitive science will generally have multiple parameters. However, we shall see, the same is also true for mathematical models in physics, and there is a standard methodology for working with such situations that we can adopt for use in cognitive science. To demonstrate this, we consider the mathematical formulation of Newton's Theory of Universal Gravitation.

$$F = \frac{G \cdot m_1 \cdot m_2}{d^2}$$

This formula tells us how to predict what force will be applied to an object by gravity, given the distance between them (d), their masses (m_1 and m_2), and the universal gravitational constant (G). Expressing this in a model-based manner, we can say that this theory lets us take a particular situation (with known masses and distances) and create a predictive model for that situation.

The important point here is that for 120 years after the development of this theory, the values of G, m_1, and m_2 *were not known*. Instead, physicists would combine the three values into a single parameter (X). Then, they would determine what value for this combined parameter best fit the particulars of a given situation.

For example, if the theory was being used to predict the influence of the Sun on Jupiter, they did not need to know the mass of either. Instead, scientists observed the path of Jupiter for a short period of time, determined the force applied by the Sun that would be required to result in such a path, and then determined for what value of X the model would give the same result. Once determined, it could then be used in all future predictions of the gravitational influence of the Sun and Jupiter. The model parameters were fit to one situation, then applied to other situations.

Just as the gravitational theory was useful for the first 120 years before the parameter G was known, so too can computational theories which do not specify the parameters of their models. In these situations, the application of the theory requires some process whereby the parameter can be determined for the situation in question. Once this process, which usually involves using some subset of the known information and finding the best-fitting parameter setting, is complete, the model can then be used to predict other aspects of the behaviour of that particular cognitive agent in that particular situation. That is, we perform parameter-fitting to create a model of this special case, and then can use that model to perform predictions. Importantly, it is this second stage which is the real test. Merely finding a parameter setting which fits does not inform us as to the veracity of the theory. By taking the further step of using the model to predict, we avoid the problems raised above by Roberts and Pashler (2000).

A more detailed theory, however, may indicate particular values for certain parameters. Once a value for *G* was included within the theory of gravitation, it could be used in new situations without the stage of first customizing the model. Well-developed theories can specify a particular value for a parameter, or can indicate a range of values, meaning that the model will be suitable no matter where in that range the parameter is set (determined by a process akin to that of sensitivity analysis). These established parameter settings become part of the theory, rather than re-fitting the values to each individual circumstance.

It is also vital to observe that *parameters need not be merely numerical values*. There is no reason why a theory might not treat an entire sub-module as a parameter. A particular component within a model might be implemented in a number of qualitatively different ways. In this case, a simple theory might say that we would have to fit the model to a given situation by finding an implementation of that component which gives a match to some aspects of the real-world behaviour (just as was true for the numeric parameters). If the resulting model can then be used to predict other aspects of its behaviour, then we have a useful theory. However, a more developed model might specify exactly what sort of implementation, or it might specify that any one of a variety of implementations could be used and still produce accurate results.

It should be noted that there is no strict distinction between numerical and non-numerical parameters. Indeed, it is always possible to build models with parameters which function as if they adjust between two different implementation systems. Furthermore, this can also occur unexpectedly (see Sibley & Kello, 2004 for an example). This means that exploring different model implementations is *as important* as exploring different parameter settings. In general, we should not be content with the current status-quo of working with one particular computational model. Instead, we need to have a variety of models (each with a variety of parameter settings). For each of these models, we can perform the equivalence testing method described previously, resulting is a numerical indication of the match between each of these models and the real world.

Generally, it is expected that we will find a large set of parameter settings which have highly similar equivalence test thresholds. The same will be true about qualitatively different models. Instead of choosing the one closest match (which will generally be a case of over-fitting), we should report what *set* of parameters and implementations result in equivalently predictive models. Further research can then discover what those successful models have in common.

Communication Techniques

In mathematical modelling, a complete representation of the model is presented within the relevant publication. The models are specified using the language of mathematics. This same language is also used to define how the models are meant to be used (i.e. how to convert a real-world situation into a model, and how to perform the statistical analysis of the resulting predictions).

However, computational models are generally significantly more complicated than can be described in a journal publication. Furthermore, we need to provide source code not only for the model, but also for the complete simulation and data analysis. To achieve parity with the mathematical approach, it should be possible to take any published paper, get access to the complete source code, run it, and have a complete replication of *all* the results and graphs from the paper.

Having such source code available simplifies the task of other researchers who wish to work with multiple sorts of models. Furthermore, having access to all of the raw data (both from the modelling results and from the real-world comparisons) resolves the problem of space limitations in publications. Within the publication, the researcher can present those aspects of the data which they believe best demonstrates the capabilities of a model, and any reader interested in other aspects can do so separately.

Modelling Software Suite

The limitations and techniques described in this paper are not new. Numerous researchers have already highlighted the problems with standard modelling approaches, and provided their own suggestions for solving them (e.g. Ohlsson, 1988; Simon & Wallach, 1999; Humphreys & O'Leary, 2002). While our approach does provide a novel justification via comparison to mathematical modelling in physics, we have extended this by developing an extensive suite of software which supports the complete modelling process. The goal is to reduce the effort required to perform the broad analysis the we recommend, involving non-standard measures, varying parameters and model types, and communicating the complete model and results.

The first toolkit is an extended library of computational models, all built to be simple to use and inter-compatible. This software suite is designed to appeal to a broad audience, and is the basis of both all of our modelling research and a graduate cognitive science course in computational modelling (Stewart, 2004). It includes Cellular Automata, Genetic Algorithms, Evolutionary Strategies, Multi-Layer Perceptrons and Back-Propagation, SRNNs, Kohonen Maps, ART, Q-Learning, and a re-implementation of ACT-R (Stewart & West, 2006). This tool set has been used by students with no previous programming background to replicate foundational modelling research, and is equally suitable for experienced programmers and researchers.

One unique feature of this software library is that it was *not* designed to be computationally efficient. Instead, the primary design criteria was simplicity and clarity of use. This results in a faster development cycle, at the expense of longer run times. We feel that the advantage of spending significantly less time developing is worth even an order of magnitude increase in required computing time. For similar reasons, the library is written in the Python programming language, which lists as one of its founding principles "Correctness and clarity before speed". For an example of the use of this system, see (West et al., 2005), where we compare mathematical, ACT-R, ANN, SRNN, and Q-Learning models of human game playing.

The second toolkit is an Web-based system that supports all of the aforementioned methodological steps we believe are required for effective modelling. The researcher writes a single program which performs a single run of their simulation (with fixed parameter settings). Internal values and final outputs are marked with a simple assignment statement, which enables the analysis software to identify the relevant data. This can be done in any programming language, or using existing modelling tools, although we focus on models developed using our model creating toolkit. For such models, the software also provides trace information about the changes in values over time during the simulation run. We find this to be a valuable debugging and interpretation tool.

Once the simulation has been run once, the source code is automatically stored. The system can be told to run the simulation multiple times, and record the results of each run (for space reasons, the changes in values over time within the model during an individual run are not recorded). Descriptive statistics, including bootstrap confidence intervals, are automatically calculated. The simulations are distributed to multiple computers running a small client program. This is suitable for running on any computer with Internet access, without requiring administrative privileges. The parameters within the model are also identified, and one can set up batch simulation runs which exhaustively vary parameter settings across given settings.

While the results of this process can be extracted for use in existing statistical tools, there is also a facility for performing the similarity-based testing described in this paper. This includes both the determination of equivalence test threshold differences between individual parameter settings and real-world results, as well as identifying parameter ranges with equivalent results. The system also produces publication-quality graphs, including contour plots detailing the effects of parameter variation.

Since the entire system is run through an Internet-based interface, the same analysis facilities provided to the researcher are also automatically provided to the community at large. By merely leaving the core interface program running (preferably on a lab server), others have complete access to the research. This includes the ability to explore the existing data set, or to download a copy of the simulation code and to generate their own investigations of the model, exploring alternate parameter settings, model changes, or alternate comparative measures.

All of this software runs equally well on any modern operating system, and all source code is released under the GNU General Public License. The complete system, including ongoing research examples, is available at <http://ccmlab.ca/ccmsuite.html>.

Acknowledgments

Funding was provided by Dr. Robert L. West's Carleton Cognitive Modelling Lab via a grant from the Natural Sciences and Engineering Research Council of Canada.

References

Anderson, J. R., Albert, M. V., Fincham, J.M. (2005) Tracing Problem Solving in Real Time: fMRI Analysis of the Subject-Paced Tower of Hanoi. Journal of Cognitive Neuroscience, 17 1261-1274.

Axelrod, R. (2005). Advancing the Art of Simulation in the Social Sciences. In Handbook of Research on Nature Inspired Computing for Economy and Management, Jean-Philippe Rennard (Ed.).Hersey, PA: Idea Group.

Beaulieu-Prévost, D. (2006). Statistical decision and falsification in science: going beyond the null hypothesis. In Hardy-Vallée, B., (ed), Cognitive Decision-Making: Empirical and Foundational Issues. Cambridge: Cambridge Scholars Press.

Connell, L. and Keane, M. (2006). A model of plausibility. Cognitive. Science, 30 (1), 95-120.

Davison, A.C. and Hinkley, D.V. (1997). Bootstrap Methods and Their Application. Cambridge University.

Erev, I. and Barron, G. (2005). On Adaptation, Maximization, and Reinforcement Learning Among Cognitive Strategies. Psych. Review, 112(4), 913-931.

Giere, R. (1988). Explaining Science: A Cognitive Approach. Chicago: University of Chicago Press

Giere, R. (1999). Science without Laws. Chicago: University of Chicago Press.

Humphrys, M. and O'Leary, C. (2002). Constructing complex minds through multiple authors. From Animals To Animats 7: (SAB-02).3-12.

Guillot, A. and Meyer, J.A. (2000). From SAB94 to SAB2000 : What's New, Animat? From Animals to Animats 6: (SAB-00). 3-12.

Ohlsson, S. (1988) Computer simulation and its impact on educational research and practice. International Journal of Educational Research, 12, 5-34.

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

Sibley, D. E. and Kello, C. T. (2004). Computational explorations of double dissociations: Modes of processing instead of components of processing. Cognitive Systems Research, 6, 61-69.

Simon, H. and Wallach, D. (1999). Cognitive modeling in perspective. Kognitionswissenschaft, 8, 1-4.

Stewart, T.C. (2004) Teaching Computational Modelling to Non-Computer Scientists. 6th International Conference on Cognitive Modelling.

Stewart, T.C. and West, R. L. (2006) Deconstructing ACT-R. 7th International Conference on Cognitive Modelling.

West, R., Stewart, T.C., Lebiere, C., and Chandrasekharan, S. (2005) Stochastic Resonance in Human Cognition: ACT-R vs Game Theory, Associative Neural Networks, Recursive Neural Networks, Q-Learning, and Humans. 27th Annual Meeting of the Cognitive Science Society.