ASSOCIATION FOR PSYCHOLOGICAL SCIENCE

Current Directions in Psychological Science
19(5) 329-335

DOI: 10.1177/0963721410386677
http://cdps.sagepub.com
SAGE

Computational Models as Aids to Better Reasoning in Psychology

Simon Farrell[1,2] and Stephan Lewandowsky[2]
[1] University of Bristol and [2] University of Western Australia

Abstract

Scientists can reason about natural systems, including the mind and brain, in many ways, with each form of reasoning being associated with its own set of limitations. The limitations on human reasoning imply that the process of reasoning about theories and communicating those theories will be error prone; we must therefore be concerned about the reproducibility of theories whose very nature is shaped by constraints on human reasoning. The problem of reproducibility can be alleviated by computational modeling, which maximizes correspondence between the actual behavior of a posited system and its behavior inferred through reasoning and increases the fidelity of communication of our theories to others.

Keywords

computational modeling, cognitive modeling, scientific reasoning, analogy, reproducibility, random walk model

Science depends on reproducibility. However much we may debate about theories, scientists tacitly assume that we are all reasoning on the same terms and from a shared understanding about data. Data that are not reproducible or that can only be reproduced under certain conditions should rightly be given little weight in reasoning about natural systems. Without this shared understanding, progress in science would be impossible, as different scientists could reach different conclusions simply because the same data are analyzed and interpreted inconsistently.

Given this emphasis on the reproducibility of experimental methods and data analysis, it is striking that little—if any—consideration is given to the fidelity and reproducibility of another core aspect of science above and beyond concerns about methodology and data analysis: scientific reasoning. Like it or not, science communication resembles a game of "telephone" wherein theories are formulated, recorded on paper, read by the next scientist who needs to understand them, and so on. Each step in this chain involves reasoning and is thus subject to known cognitive limitations. Numerous experiments have established those limitations, and scientific reasoning is indubitably not exempt from them.

Limitations on Thinking

A worryingly long list of limitations on human thinking was presented by Hintzman (1991), who argued that aspects of cognition such as the *confirmation bias* (see, e.g., Evans, 1989) and the limited capacity of working memory (e.g., Engle & Kane, 2004) have profoundly negative implications for scientific reasoning. Equally worryingly, it is easy to extend Hintzman's list. For example, when given the sequence "2-4-6," people tend to develop overly specific and baroque theories of the generating rule, such as "numbers increasing in units of 2" (see, e.g., Evans, 1989); few discover the experimenter's rule of "any increasing series." This finding arguably has parallels in science: For example, putatively specific deficits in grammar may instead reflect more general cognitive deficits (e.g., Christiansen & Ellefson, 2002).

Similarly, scientists often draw *analogies* between a source domain (in which the interrelations between elements are known) and a target domain (where the relationships are unknown). Unfortunately, analogies can be misapplied. Gentner and Gentner (1983) identified two common analogies to understand electricity: water flowing through pipes, and crowds of people running through passageways. Gentner and Gentner found that errors made on electrical circuit problems depended on which analogy an individual adopted. This problem of analogical reasoning has implications for science, where relying on analogies could potentially produce misunderstanding of a psychological system. As a case in point, consider the popular spreading-activation theory, which postulates

Corresponding Author:
Simon Farrell, Department of Experimental Psychology, University of Bristol, 12a Priory Road, Clifton Bristol BS8 1TU, United Kingdom
E-mail: simon.farrell@bristol.ac.uk

Table 1. Problems Affecting Reasoning and Shared Understanding Among Scientists and Ways in Which Computational Modeling Can Address Them

Reasoning problem	Solution from computational modeling
Reasoning	
Confirmatory bias: tendency to seek out evidence that confirms (rather than disconfirms) a hypothesis	Emphasis on comparing multiple models; model selection allows us to find evidence for and against models
Analogical reasoning: reliance on different analogues; contents of source domain leak in to inference about target domain	Formal system means theory will behave the same way regardless of differences in analogies adopted by individual scientists
Reasoning about complicated, distributed or massively parallel structures: restricted by working memory limitations and bias to interpret such networks unidirectionally (e.g., White, 2008)	Models not limited by working memory limitations; many models are highly distributed or complicated in nature (e.g., connectionist models)
Incompleteness of reasoning: there can be more in the data than might be inferred from standard analyses of performance (e.g., proportion correct)	Fitting a model can reveal hidden structure or processes (emergent phenomena) that are not directly inferable from standard analyses of performance
Shared Understanding	
Precise communication	Computer code (commented to make links with textual description) can be shared between researchers in a similar fashion to data and analysis files
"That's not what I meant" problem: Shared understanding of theories, and falsifiability of theories, difficult to achieve when specification of theory is fuzzy	With clear, computationally formulated definition, theories make unambiguous predictions (at least within target domain) and are more falsifiable.

Note: For examples of other constraints on reasoning and the solutions offered by modeling, see Hintzman (1991) and Lewandowsky (1993).

that concepts are represented by an interconnected network of nodes. Nodes are activated upon stimulus presentation, and activation spreads through the connections to neighboring nodes. To understand and communicate the notion of spreading activation, several analogies might be used—for example, electricity passing through wires (e.g., Radvansky, 2006) or water passing through pipes. Those analogies help determine our understanding of the model's behavior. The water analogy necessarily implies a relatively slow spread of activation, contrary to the data, which imply activation of distal concepts is almost instant (Ratcliff & McKoon, 1981). Conversely, the electricity analogy can handle the instantaneity but places the explanatory burden on the links between nodes—akin to circuits being closed—rather than on their activation. Although analogies are often drawn within a scientific domain (e.g., "this data pattern resembles one I've seen recently"; e.g., Dunbar & Fugelsang, 2005), many psychological mechanisms have no known source domain in psychology, and theorists will therefore be required to draw analogies to systems outside psychology in which they—and other psychologists—may not be experts.

A second problem, related to these vagaries of reasoning, is that the shared understanding of a system by multiple individuals depends on the extent to which all involved reason identically. A group of scientists can have a shared (though perhaps flawed) understanding of a system if all scientists adopt the same analogy. If analogies are heterogeneous, mutual misunderstandings are inevitable: Two scientists may reach mutually incompatible hypotheses about spreading activation if they adopt different analogies (water vs. electricity). Furthermore, scientists may use any number of modes of thinking or representations to conceptualize a theory and to derive predictions (Markman & Gentner, 2001): Scientific reasoning may be accomplished by constructing mental models (Gentner & Gentner, 1983), by using visual representations such as sketches and graphs (Oestermeier & Hesse, 2000), and by mental simulation (Trickett & Trafton, 2007). Not only are these different modes imperfect, but there is also no guarantee that any two scientists will use the same mode of reasoning and thus share an understanding of a system.

In light of these concerns and others (see left column of Table 1), do we really have a deep understanding of the theories we reason about? Do we really know what our theories predict? Do our colleagues understand our theories the same way we do?

Computational Models as an Aid to Reasoning

Those reasoning problems can be alleviated by implementing one's theoretical principles as a computational model (or equations in a mathematical model). A principal advantage of computational modeling is that we are forced to specify *all* parts of our theory. In the case of spreading activation, we must answer such questions as Can activation flow backward to immediately preceding nodes? Is the amount of activation unlimited? Is there any leakage of activation from nodes? These further specifications, which verbal theories omit altogether, render our theory more readily communicable and more falsifiable. A summary of such advantages of computational modeling is provided in the right-hand column of Table 1.

Computational models thus check whether our intuitions about the behavior of a theorized system match what actually arises from its realization. To illustrate, consider the "random

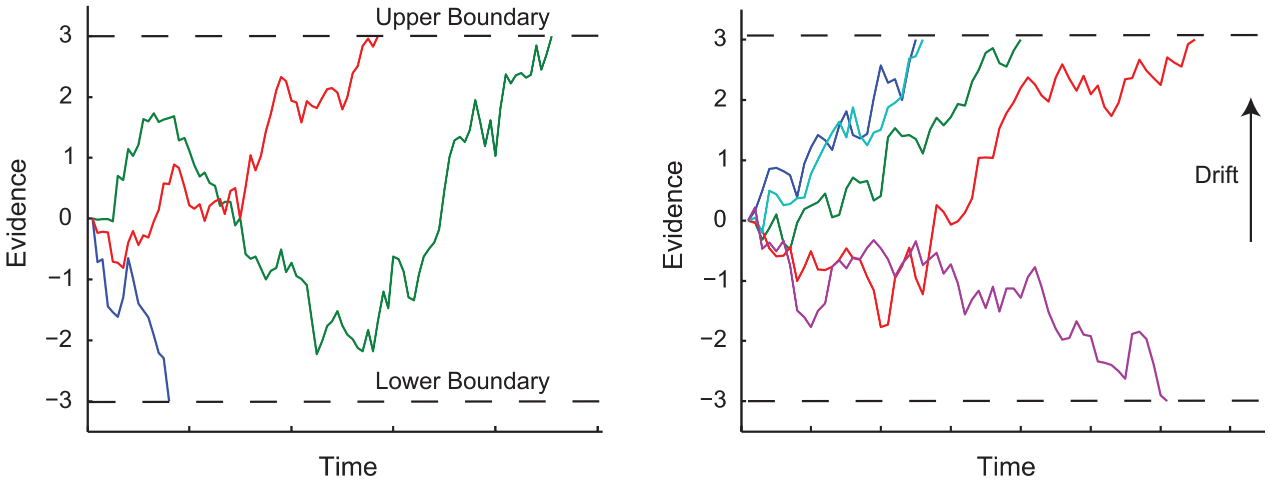

Fig. 1. The random-walk model of decision making. At each time step, evidence is sampled from an information source, and the evidence is accumulated over time. When the information exceeds one of the two decision boundaries (upper or lower), the respective decision is made. The left panel shows the summed evidence for three different trials. The three paths all rely on the same source of information and differ only in the random sampling of the information over time. The red and green paths have reached the upper boundary and would indicate one decision (e.g., "The tomato is green"), while the blue path has reached the bottom boundary corresponding to the other decision (e.g., "The tomato is red"). The right panel shows five sample paths for a case in which there is a tendency to drift towars the upper boundary because the information favors that decision on average.

walk" model of binary decisions, in which people are assumed to sample evidence from their environment in discrete steps and then sum this sampled evidence to make a decision between two alternatives (for example, deciding whether a tomato is ripe enough to eat based on its color). At each step, a sample can nudge the summed evidence toward one decision or another, with the size and direction of the nudge being random, giving the random-walk model its name. We can simulate a random walk ourselves by standing in the middle of a hallway, repeatedly tossing a coin, and taking a step up the hallway if the coin comes up heads or down the hallway if the coin comes up tails. Samples of evidence continue to be taken until a threshold—for example, either of two lines drawn across the carpet—is crossed. At this point, depending on which of two thresholds the evidence trail crosses, one decision (e.g., "The tomato is green") or the other ("The tomato is red") is reached. The left panel of Figure 1 shows some illustrative random walks for the case in which the information is equally favorable to the two alternatives (for example, where the crossing light is covered up, corresponding to the flip of an unbiased coin in our hallway analogy). Each path shows how the summed evidence changes as additional samples are taken, with the paths terminating when one of the two response thresholds (the dashed lines in the figure) is reached. The paths are different because of the randomness in the random walks: In the blue path in the left panel of Figure 1, a number of negative samples happen to have been drawn at the start, immediately pushing the random walk toward the bottom boundary.

One feature of the random-walk model that makes it ideal for modeling decision behavior is that it predicts both the probability and the time taken to make a decision, this time being equivalent to the number of steps that were taken before one of the two thresholds was reached. In cases in which there is equal evidence for the two alternatives (as is the case for the random walks in the left panel of Fig. 1), the probabilities of the two decisions are equal, and they are predicted to have identical response-time characteristics. Now imagine how the model will behave when the evidence favors one decision over the other (as would be the case when our tomato is a rich red and ready to eat). This introduces some "drift" toward the favored threshold by "biasing" the sampled information—in our hallway analogy, this would correspond, for example, to the coin being strongly biased to come up heads. This bias is illustrated with some example random walks in the right panel of Figure 1. As in the left panel, the evidence sampled on each step is random, but is more likely to be a positive number, meaning the random walk is more likely to take a step upward rather than downward. What do you expect will happen to the probability of making one choice over another? What about the time taken to make each choice? Under these circumstances, the drift will increase the probability of the evidence crossing the upper boundary (as you can see in the right panel, where four out of the five random walks hit the top threshold). One might also predict that the response time would be slower for the less likely response. The latter prediction is actually incorrect; in fact, the mean response times for the two alternatives are the same, as can be seen by comparing the upper and lower histograms in Figure 2. The bars represent ranges of response times; the top panel represents cases in which the random walk hits the top threshold (corresponding to the favored response) and the bottom panel represents cases in which it hits the bottom threshold (less favored response). The two histograms are virtually identical (Stone, 1960).

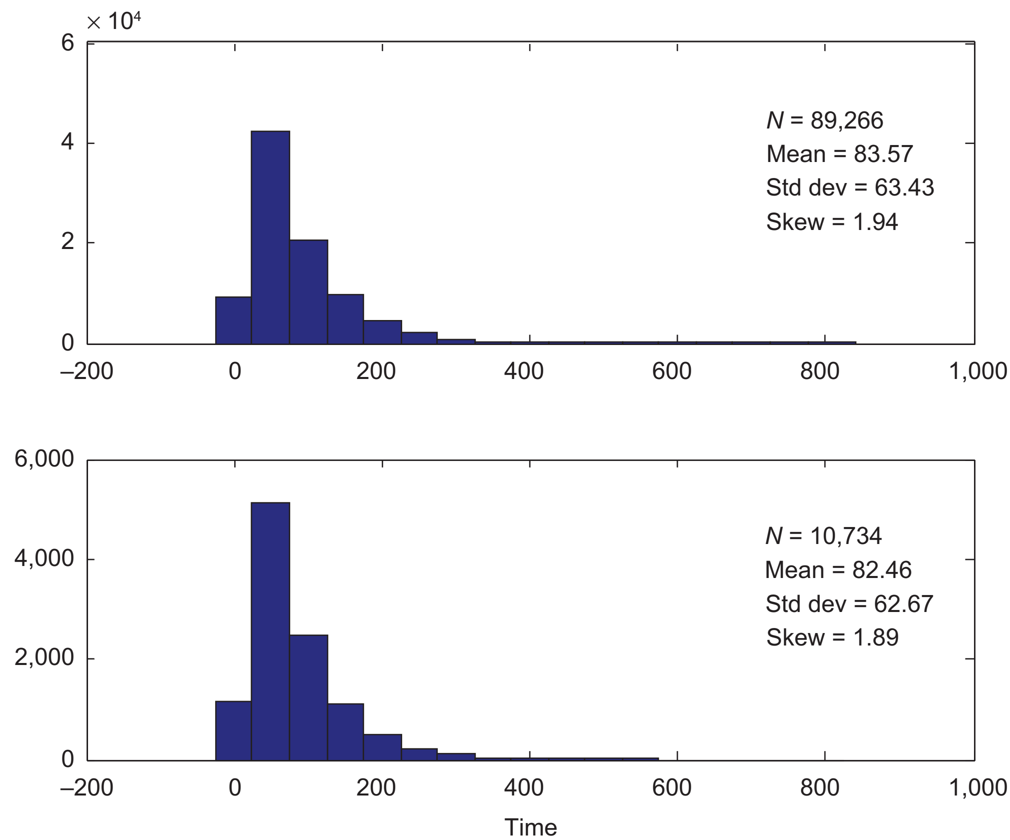

Fig. 2. Histograms of decision times for decisions crossing the upper boundary in the right panel of Figure 1 (top panel) and lower boundary (bottom panel) from 100,000 simulation runs of the random-walk model. Although the upper-boundary decision is more likely to be made (around 89% of the time, indicated by *N*), the two latency histograms have an identical appearance, with the same mean, standard deviation, and skew (the slight differences are due to the randomness in the simulations).

This seems a little strange—after all, doesn't the upward drift mean that it will take longer for a random walk to reach the bottom boundary, like a person struggling against a river current? Or maybe you (like us) pictured "rays" emanating from the starting point representing some reasonable range of average trends, and imagined this rotating upward in the case where some drift is introduced to produce slow "bottom" responses (see Fig. 3). The swimmer analogy misses the important detail that the only systematic pressure is the drift (unlike the swimmer, who by definition is applying his or her own "counter-drift" against the current). This means that paths that hit the bottom boundary do so only by the chance happenstance of having collected a series of samples that work against the drift. This also explains why the "ray" analogy in Figure 3 fails—having a slower rate of approach to the bottom boundary slows those responses down in the analogy, but in reality, any additional time gives those paths more time to be bumped toward the top boundary. We can also reject this analogy by a logical consideration, as follows. In the ray model, if we take less time to reach a certain level of evidence (for example, if the summed evidence adds up to –2, or two units of evidence below the starting point), we will take less further time to hit the bottom boundary compared to the case in which we have taken longer to reach the same intermediate point. This can be seen by considering the right-pointing horizontal arrow drawn through the two bottom paths (that happens to mark a summed evidence value of –2) and noting that the lines diverge. However, the random-walk model is by definition agnostic regarding the time that has already passed; if we have reached a summed evidence of –2, we have a constant probability of taking a downward (vs. an upward) step. The behavior of the basic random-walk model is not at all obvious from its description and shows up the limitations on our reasoning about such processes.[1]

By revealing the "real" behavior of a system, modeling can generate insights that conventional reasoning processes may fail to uncover. For example, in the random-walk model, we can differentiate the drift rate (the quality of information) from the separation between the boundaries (the amount of evidence needed to make a decision, called boundary separation). Schmiedek, Oberauer, Wilhelm, Süß, and Wittmann (2007) did exactly that by fitting a variant of the random-walk model to data from several choice-reaction-time tasks, and found that scores on intelligence tests were more strongly related to drift rate (i.e., the strength of the bias during sampling of the information in the random walk) than to boundary separation (i.e., how far the two thresholds are from the origin of the

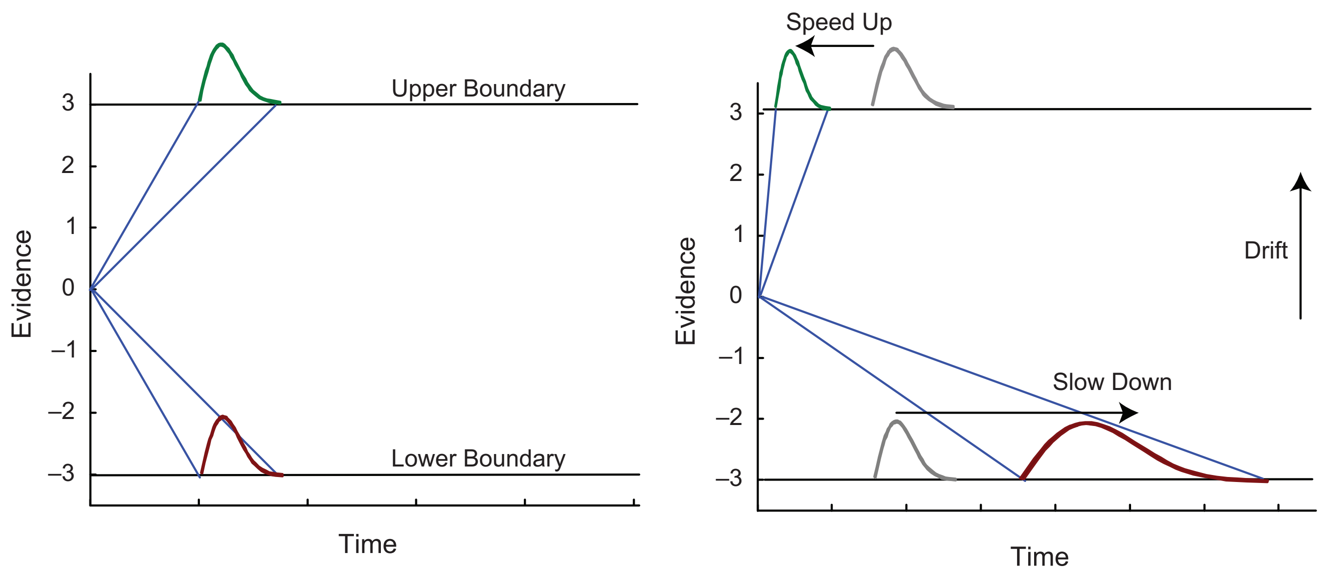

Fig. 3. A schematic depiction of an intuitively reasonable but *incorrect* mental simulation of the effects of drift in the random-walk model. The left panel depicts some representative rays emanating from the starting point in the situation in which drift = 0. When drift toward the positive (top) boundary is introduced, the rays might incorrectly be assumed to rotate in that direction, producing a speeding of top responses and a slowing of bottom responses.

random walk), suggesting that information extraction is a fundamental aspect of intelligence. Moreover, we frequently observe behavior emerging from a model in which no such behavior is specified by the programmer at the outset. Elman (1990) found that when he exposed a particular type of model, called a recurrent network, to the sequential statistics in a large corpus of natural language, the model spontaneously formed identifiable representations of nouns, verbs, and adjectives, even though these grammatical types were not specified in the input or in the model itself.

Finally, computational modeling helps ensure reproducibility in scientific thinking. By implementing a model as a computer program or a set of equations, another researcher can take our model and exactly reproduce our predictions. For example, you may still doubt whether the random-walk model really produces the same latency distributions for the different decisions in the presence of drift. With a little experience in a computer language such as R or MATLAB, you would be able to simulate the model as described and confirm that prediction for yourself. As an additional step, publishing the model code on the Web facilitates testing and exploration by others. Use of the same formalized description ensures that we ultimately derive the same predictions from a theory and assists in forming a shared conceptual understanding.

To illustrate the last point, consider the concept of inhibition that is often invoked to explain differences between individuals in their ability to perform a task (see MacLeod, Dodd, Sheard, Wilson, & Bibi, 2003, for a review). For example, the phenomenon of negative priming—whereby a response to a stimulus is slowed when that stimulus recently appeared as distracting information—is taken as evidence for an inhibitory component of selective attention (Tipper, 1985). However, MacLeod and colleagues suggested that "inhibition" is a vague term that often amounts to little more than the renaming of an observed difference between two mean latencies, giving fertile ground for confusion or lack of shared understanding between researchers. This problem is overcome within a computational model, such as the choice model of Brown and Heathcote (2005), in which representations of response alternatives compete for selection. The model provides several mechanisms by which inhibition can occur, including (a) reduced external input to one or more alternatives; (b) increased "leakage" from alternatives; (c) a reduction in the resting baseline of an alternative; and (d) a tendency for more active units to reduce the activation of other units, called "lateral inhibition." By specifying the operation of the model (see, e.g., Equation 1 of Brown & Heathcote, 2005), the source of inhibition is made explicit, leaving other researchers (and ourselves) with a better grasp of the meaning of "inhibition."

Current Directions

Contemporary theorizing increasingly involves quantitative comparison of competing models to weigh the evidence for and against various theories in light of a particular data set. Whenever a model provides a better quantitative explanation of the data than other models, it receives further support. These comparisons can additionally be corrected for the complexity of a model. One widely accepted dictate in science is "Occam's Razor": We should prefer the simplest theory that adequately explains the data. Determining the proper level of complication of a theory is sometimes difficult, and techniques for the quantification of complexity continue to be developed (Pitt & Myung, 2002). Another advance is the incorporation of psychological models into common statistical frameworks. For example, individual-differences research often employs structural equation modeling, in which relationships between observed variables are captured via "latent" variables that represent psychologically meaningful constructs (e.g., working memory capacity). Recent research has used psychologically meaningful model parameters (e.g., the strength of the drift of a random walk) as variables that can be used in structural equation models (Schmiedek et al., 2007). Furthermore, computational modeling can inform research from cognitive neuroscience, where model parameters can be linked to brain activity (as recorded by electroencephalography or functional magnetic resonance imaging) and can be used to make inferences about changes in cognitive mechanisms, inferences that could not otherwise be made (e.g., Ho, Brown, & Serences, 2009).

In closing, we should note that modeling is not a panacea for all scientific ills. In some situations, the extent to which an algorithm or equation is mandated by (and itself mandates) a psychological interpretation may be questionable. For example, much work has been dedicated to comparing different mathematical functions in their ability to account for the extent of forgetting over time (e.g., Rubin & Wenzel, 1996); however, arguably there is a limit on how much these functions can tell us in the absence of a model of the underlying mechanisms. Bearing in mind this caveat, computational modeling should have a place in any psychologist's toolkit—alongside experimental design and statistics—as a way of developing, understanding, and communicating theories.

Recommended Reading

Carruthers, P., Stich, S., & Siegal, M. (2002). *The cognitive basis of science.* Cambridge, England: Cambridge University Press. An edited volume that provides a good overview of cognitive approach to scientific reasoning and the social and epistemic context in which scientific reasoning takes place.

Hintzman, D.L. (1991). (See References). A classic chapter arguing for the widespread formal use of computational models in psychology.

Lewandowsky, S., & Farrell, S. (in press). *Computational modeling in cognition: Principles and practice*. Thousand Oaks, CA: Sage. A new textbook that explains the logic behind computational modeling and works through the steps of developing and testing models in cognitive psychology.

Notes

1. In fact, other types of evidence-summing models do predict different latencies for the two response classes, but they do this by effectively making the random-walk model more like the ray analogy.

Acknowledgments

We thank the anonymous reviewers for their helpful comments.

Declaration of Conflicting Interests

The authors declared that they had no conflicts of interest with respect to their authorship or the publication of this article.

Funding

Preparation of this paper was facilitated by a Discovery Grant from the Australian Research Council and an Australian Professorial Fellowship to the second author and a Linkage International grant from the Australian Research Council to both authors.

References

Brown, S., & Heathcote, A. (2005). A ballistic model of choice response time. *Psychological Review*, *112*, 117–128.

Christiansen, M.H., & Ellefson, M.R. (2002). Linguistic adaptation without linguistic constraints: The role of sequential learning in language evolution. In A. Wray (Ed.), *Transitions to language* (pp. 335–358). Oxford, England: Oxford University Press.

Dunbar, K., & Fugelsang, J. (2005). Scientific thinking and reasoning. In K.J. Holyoak & R. Morrison (Eds.). *Cambridge handbook of thinking and reasoning* (pp. 705–726). Cambridge, England: Cambridge University Press.

Elman, J.L. (1990). Finding structure in time. *Cognitive Science*, *14*, 179–211.

Engle, R.W., & Kane, M.J. (2004). Executive attention, working memory capacity, and a two-factor theory of cognitive control. In B.H. Ross (Ed.). *The psychology of learning and motivation.* (Vol. 44, pp. 145–199). New York, NY: Elsevier.

Evans, J.S.B.T. (1989). *Bias in human reasoning: Causes and consequences.* Hove, England: Erlbaum.

Gentner, D., & Gentner, D.R. (1983). Flowing waters or teeming crowds: Mental models of electricity. In D. Gentner & A.L. Stevens (Eds.), *Mental models* (pp. 99–129). Hillsdale, NJ: Erlbaum.

Hintzman, D.L. (1991). Why are formal models useful in psychology? In W.E. Hockley & S. Lewandowsky (Eds.). *Relating theory and data: Essays on human memory in honor of Bennet B. Murdock.* (pp. 39–56). Hillsdale, NJ: Erlbaum.

Ho, T.C., Brown, S., & Serences, J.T. (2009). Domain general mechanisms of perceptual decision making in human cortex. *Journal of Neuroscience*, *29*, 8675–8687.

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

MacLeod, C.M., Dodd, M.D., Sheard, E.D., Wilson, D.E., & Bibi, U. (2003). In opposition to inhibition. In B.H. Ross (Ed.), *The psychology of learning and motivation* (Vol. 43, pp. 163–214). San Diego, CA: Elsevier.

Markman, A.B., & Gentner, D. (2001). Thinking. *Annual Review of Psychology*, *52*, 223–247.

Oestermeier, U., & Hesse, F.W. (2000). Verbal and visual causal arguments. *Cognition*, *75*, 65–104.

Pitt, M.A., & Myung, I.J. (2002). When a good fit can be bad. *Trends in Cognitive Sciences*, *6*, 421–425.

Radvansky, G. (2006). *Human memory*. Boston, MA: Pearson.

Ratcliff, R., & McKoon, G. (1981). Does activation really spread? *Psychological Review*, *88*, 454–462.

Rubin, D.C., & Wenzel, A.E. (1996). One hundred years of forgetting: A quantitative description of retention. *Psychological Review*, *103*, 734–760.

Schmiedek, F., Oberauer, K., Wilhelm, O., Süß, H.-M., & Wittmann, W.W. (2007). Individual differences in components of reaction time distributions and their relations to working memory and intelligence. *Journal of Experimental Psychology: General, 136*, 414–429.

Stone, M. (1960). Models for choice-reaction time. *Psychometrika*, *25*, 251–260.

Tipper, S.P. (1985). The negative priming effect: Inhibitory priming by ignored objects. *Quarterly Journal of Experimental Psychology, 37,* 571–590.

Trickett, S.B., & Trafton, J.G. (2007). "What if . . . ": The use of conceptual simulations in scientific reasoning. *Cognitive Science, 31*, 843–875.

White, P.A. (2008). Beliefs about interactions between factors in the natural environment: A causal network study. *Applied Cognitive Psychology*, *22*, 559–572.