The Journal of Mind and Behavior
Autumn 2009, Volume 30, Number 4
Pages 337–344
ISSN 0271–0137

Cambridge Handbook of Computational Psychology. Ron Sun [Editor]. New York: Cambridge University Press, 2008, 753 pages. $65.00 soft cover.

Reviewed by Robert L. West, Carleton University

Computational psychology refers to the effort to create computational mechanisms that, in some way, mimic mechanisms within the brain. More specifically, the goal in creating these mechanisms is to show that they can systematically reproduce patterns of human behaviour elicited under specific conditions. From this it is inferred that these mechanisms bare some similarity to the brain mechanisms that produced the human behaviours. In most cases this involves mimicking the results of psychology experiments, although it is good to see in this book, two chapters discussing the application of this approach to non experimental areas (multi agent social interactions and cognitive engineering).

The *Cambridge Handbook of Computational Psychology* is a well-organized book. It begins in Part I with an introductory chapter by the Editor, Ron Sun. In Part II, "Cognitive Modeling Paradigms," various cognitive modeling paradigms are introduced in separate chapters. The paradigms are: connectionist models, Bayesian models, dynamic systems, logic, and cognitive architectures. The chapters in this section give an overview of how modeling occurs within the paradigm and discuss the ideas and perspectives associated with the paradigm. These chapters make broad arguments about the usefulness of each paradigm and provide specific examples. This section is followed by Part III, "Computational Modeling of Various Cognitive Functionalities and Domains," in which each chapter describes the types of models used within specific research areas. The research areas are: episodic memory, semantic memory, categorization, decision making, inductive reasoning, mental logic, skill acquisition, implicit learning, attention and cognitive control, developmental psychology, psycholinguistics, personality and social psychology, social simulation, scientific explanation, cognitive engineering, animal learning, vision, and motor control. The chapters in this section mainly provide an overview and discussion of the specific cognitive models used in each research area. This section is followed by Part IV, "Concluding Remarks," which contains a chapter called, "An Evaluation of Computational Modeling in Cognitive Science," by Margaret A. Boden, which focuses on understanding the meaning of cognitive modeling, and a chapter called, "Putting the Pieces Together Again," by Aaron Sloman, which looks at prospects for the future of cognitive modeling.

Requests for reprints should be sent to Robert West, Ph.D., Department of Psychology, Carleton University, 1125 Colonel By Drive, Ottawa, Ontario, Canada, K1S 5B6. Email: robert_west@carleton.ca

Overall, this format makes for a very useful reference book. However, books such as this cannot be written both for the expert and for the novice, and this book is written more for the expert. In particular, this is an excellent book for anyone with expertise in a particular area of modeling, who wants to learn more about other methods. This book is potentially useful for someone who wishes to better understand cognitive modeling, but does not have a background in this area. However, for the uninitiated, the book will make more sense with a bit of background information.

The terms, computational psychology, cognitive modeling, and cognitive architectures are often used interchangeably. However, defining some differences between these terms will make the goals and purposes of this book clearer. For purposes of understanding this book, computational psychology can be defined as the study of how computational mechanisms can function as descriptions of brain mechanisms. Cognitive modeling is the business of building models that use these mechanisms to mimic human behaviours. If the application of a mechanism to a behaviour is very direct then there is very little difference between computation psychology and cognitive modeling. However, this is frequently not the case, so cognitive modeling often represents specific ways of getting mechanisms to mimic human behaviours. This would include different ways of setting up the mechanisms as well as various additional assumptions needed to get it to work (one of the benefits of building cognitive models is that it makes these additional assumptions explicit). Cognitive architectures are much more ambitious structures, aimed not at specific classes of behaviours, but at a broad range of human abilities. Cognitive architectures are intended to explain how different mechanisms can represent different human abilities, *and* how these mechanisms interact to produce complex behaviours. The mechanisms within an architecture can all come from the same family (e.g., they may all be neural networks), or the architecture can be a hybrid, combining mechanisms from different families (e.g., Sun's Clarion architecture combines neural networks with symbolic systems; for a discussion see Chapter 1 in the book). It is also important to note that there are currently no cognitive architectures that claim to be able to model all mental functions. Some are more ambitious than others in terms of their intended scope, but the point is that a model becomes an architecture when it involves combining mechanisms to produce complex and varied behaviours.

The meaning of these approaches for the study of the mind is discussed in the introductory and concluding chapters. Three points are important to note for people unfamiliar with this area. The first is that computational models are viewed as a type of theory. Some people do not believe that a computational model of a phenomenon constitutes a theory about the phenomenon, but this point of view, while acknowledged, is not given much consideration (see Chapter 1 by Ron Sun for a discussion). The second point is that cognitive modeling is viewed as *essential* for the field of psychology to move ahead in a scientific way. Again, not much consideration is given to contrary points of view. However, quite a bit of discussion *is* given to finding the best way to move forward so that cognitive modeling can achieve its full potential (see Chapter 1 and Chapter 26 by Aaron Sloman for discussion). The third point is that cognitive modeling is viewed as being potentially able to account for all mental functions, including such things as emotions and consciousness. This issue causes confusion because cognitive psychology has historically focused on *cold cognition*, e.g., memory, problem solving, etc., whereas the field of cognitive science is based around understanding all brain/behaviour functions from an information processing view. One way of viewing it is that the cognitive revolution began by studying things that had a more obvious relationship to the way that computers work, but the intention has always been to explain all mental functions (see Chapter 25 by Margaret Boden for a discussion of this issue, including brief mention of dissenting views). Essentially, this is a positive book about cognitive modeling; it represents the views of people who believe in this endeavour. It is also a relatively up to date book, which is important because the field of cognitive modeling has evolved significantly and a lot of the criticisms and issues historically associated with this area are now out of date.

One issue that could potentially cause confusion is the role of cognitive architectures in the book. Cognitive architectures are discussed in the introductory chapter and the concluding chapters. They are also discussed in Chapter 6 by Niels Taatgen and John Anderson. While this chapter compares several different architectures, it does not explain them in detail. Instead the focus is on how to scientifically evaluate cognitive architectures. This is a bit confusing because it does not explain a paradigm, but rather how to test models created within a paradigm. However, as noted above, cognitive models are mechanisms; cognitive architectures are combinations of mechanisms. Since mechanisms can potentially be combined in many different ways it is not really possible to describe how "cognitive architectures" work. So, this chapter, while important, is a bit different from the other chapters in this section. As for Part III of the book, "Computational Modeling of Various Cognitive Functionalities and Domains," some chapters discuss the use of cognitive architectures, in particular Chapter 13, "Computational Models of Skill Acquisition," by Stellan Ohlsson, and Chapter 21, "Cognitive Modeling for Cognitive Engineering," by Wayne Gray, provide good discussions of cognitive architectures, but again, they do not explain the architectures in detail. So, although cognitive architectures play a prominent role in this book, this is not a book for learning about specific architectures (see Sun, 2006 for reviews of some well known architectures). However, it is not possible to have a full discussion about computational psychology without considering the role of cognitive architectures.

The idea of cognitive architectures was proposed by Alan Newell to deal with the problem that psychological theories (including cognitive models, which will be assumed to constitute valid psychological theories for the remainder of this review) only explain very narrow sets of experimental findings (see Newell, 1990 for a full discussion). This results in two related problems, (1) the theories (including computational models) do not explain how different aspects of cognition interact to produce an integrated, intelligent system, and (2) the theories (including computational models) are underconstrained, precisely because they are so narrow. Newell's answer was to commit to broad theories that make detailed predictions, covering a range of human abilities. He called these theories, cognitive architectures, and envisioned that they would be sufficiently complex to require the use of computer simulations to work with them (i.e., similar to complex physics models or environmental models, such as climate change).

The biggest criticism of computational psychology is that the models can be made to "fit" any data set by customizing the model and/or manipulating the parameters (Roberts and Pashler, 2000 are most frequently referenced in support of this). The idea of cognitive architectures deals with this problem by requiring a single mechanism to explain all the behaviours related to the cognitive ability represented by the mechanism. Importantly, this includes behaviours that are produced through the integrated co-ordination of different cognitive abilities. Taking this approach puts many more constraints on theories and also makes the extent to which a model has been "fitted" to the data, fairly obvious. Specifically, the legitimacy of a "fit" can be judged by the extent to which the mechanism (including the parameter values) had to be altered from pre-existing successful applications of the mechanism to other behaviours. The Taatgen and Anderson chapter pursues this theme further, examining different principles that can be applied to make cognitive architectures into useful and testable theories. This is why this is an important chapter.

Also, as discussed in Aaron Sloman's concluding chapter, the goal of building integrated models of human cognition forces us to confront issues that can be neatly avoided by studying narrow areas in isolation. Sloman uses the concept of "scaling out," instead of "scaling up" to point out that humans use the same mechanisms for very different purposes and often combine mechanisms to produce different results. However, Sloman's point is not about improving the testability of cognitive models. Instead, his point is that thinking in this way will lead us to new insights that cannot be achieved using the traditional divide and conquer, reductionist approach of experimental psychology. However, the argument is not that traditional experimental psychology should be discarded, but rather that it must be augmented.

In this book, a discussion of cognitive architectures arises whenever larger theoretical issues are discussed. This is interesting because it suggests that the idea of cognitive architectures has become central to the field of cognitive modeling, which was not always the case. Although most of the models described in this book are not built within a cognitive architecture, it is clear that one-off models of narrow data sets are not seen as the way forward. Overall, the picture that arises from the book is that most research still uses individual mechanisms to model specific phenomena, but that this is not the end goal. The book suggests a growing consensus that the goal is to integrate these models into larger architectures, which will represent well constrained scientific theories (see the chapter by Taatgen and Anderson) and tools for gaining new theoretical insights into the way the mind works (see the chapter by Sloman).

This book also reflects a new, co-operative ethos in cognitive modeling. Cognitive modeling is no longer viewed as a battle to determine which mechanism represents the right way of understanding the human mind (e.g., neural networks versus symbolic systems). Instead, the new way lies in understanding the relative strengths and weaknesses of different mechanisms for modeling different aspects of human cognition. As Sloman notes, we need to understand the space of possible mechanisms that can be applied to a problem. In Chapter 25, Margaret Boden provides a brief history of cognitive modeling, which includes references to various past struggles between paradigms. However, that sense of competition, that was still common only a decade ago, is almost entirely absent in this book. Instead, what comes across is a need and a desire to build a common framework and methodology for developing and testing models. In support of this, a number of chapters make reference to hybrid architectures (i.e., combining the strengths of different mechanisms) and to the fact that the functionality of some mechanisms can be obtained using other mechanisms, and that this is a good thing because it leads to insight (e.g., see Chapter 2 by Michael Thomas and James McClelland for a discussion of the relationship between connectionist networks and Bayesian systems).

That being said, it is also clear that this is a very diverse field. The organizational structure of the book reflects this. Cognitive modeling is divided among people working within fundamentally different paradigms (Part II, "Cognitive Modeling Paradigms"). Likewise, the application of modeling paradigms is divided among people interested in very different mental abilities (Part III, "Computational Modeling of Various Cognitive Functionalities and Domains"). Furthermore, the modeling paradigms and the fields of study in which they are applied are both very complex, requiring years of study. Therefore, it is extremely difficult to be fully knowledgeable in all areas of cognitive modeling. To investigate how this diversity affects the integration of ideas I did a case study, looking at how my own research relates to different parts of the book. One of my areas of research is on human game playing. Specifically, I and my colleagues use the ACT–R cognitive architecture (Anderson and Lebiere, 1998) to model how humans interact in simple game playing situations (e.g., see West, Lebiere, and Bothell, 2006). Game playing is covered in the book in Chapter 10, "Micro-Process Models of Decision Making," by Jerome Busemeyer and Joseph Johnson, but our research is not reported there. Instead, it is reported in Chapter 19, "Cognitive Social Simulation," by Ron Sun. Why? I think it is because the methodology we use focuses on having two intelligent players (humans and/or ACT–R agents) interact, so it is a type of social interaction and we study it by simulating it. In contrast, most research on game playing involves studying various aspects of game playing in isolation (e.g., how well humans can behave randomly or how well humans can learn payoff schemes). So although both methods can be applied to game playing, they are very different research paradigms. Another difference is that our research is focused on how the human declarative memory system is used to predict what someone will do next. So our work is viewed more as research on the ACT–R model of declarative memory than on game playing per se. In contrast, game playing has traditionally been studied within the game theory paradigm. Although, as described in the Chapter 10, by Jerome Busemeyer and Joseph Johnson, psychological models deviate from the game theory paradigm, they are still mainly understood in terms of how they deviate from that paradigm. Overall, I think this ambiguity over where our model belongs is pretty typical and reflects the difficulty in bringing together people using different methodologies, different modeling paradigms, and working in different research areas.

It was also interesting to note that in Boden's review chapter she reports Clarion (see Chapter 1 by Sun for a brief description) as the only architecture designed to model social interactions. This is not surprising since ACT–R is not generally used in this way but, in fact, any architecture can be adapted for multi agent modeling (although it might not be easy) so in this sense it can be said that certain architectures are designed for multi agent interactions. In fact, we used Python ACT–R (Stewart and West, 2006), which is a version of ACT–R (implemented in Python) that is designed for modeling multi agent interactions. This highlights another issue, which is that there are a lot of systems and versions out there that are not well known. In addition, they are not very well documented, so even if you can find them, they are still hard to understand. Another thing that can be seen by going through the chapters in Part III is that the modeling paradigms used in different research areas are not necessarily chosen because of some sort of match between the paradigm and the phenomena. Instead, the use of a particular paradigm in a particular area is often historic; propelled forward by previous work in the area. One way in which this book could be particularly useful is to make people aware of other choices and possibilities.

To get an idea of the potential number of connections that could be generated through the use of this book, I again used the work I have done with my colleagues on games as a case study, by asking how many chapters are relevant to our work. Since we are doing cognitive modeling, the introductory and concluding chapters on cognitive modeling are obviously relevant. Also, we use the ACT–R architecture, so the chapter on cognitive architectures and how to make them testable is relevant. However, we have also implemented our model as a neural network, so the neural network chapter is relevant, and since neural networks can be interpreted in Bayesian terms, this means the model could be implemented in a Bayesian form, so the Bayesian chapter is also relevant. The chapter on dynamic systems is relevant because we use a dynamic systems framework to understand and explain our results (e.g., randomness is generated through a "chaotic" process). The chapter on decision making is relevant because that is where game playing has traditionally been studied, and the chapters on episodic and semantic memory are relevant since we use the ACT–R declarative memory, which is currently used for modeling both episodic and semantic memory (and it's not clear whether our model should be classified as episodic or semantic). The chapters on inductive reasoning, deductive reasoning, skill acquisition and implicit learning are all relevant because they discuss modeling methods that could be applied to game playing, and therefore constitute potential alternatives to our model (in fact, we have compared our model to recurrent networks, which are discussed in the implicit learning chapter, and procedural learning systems, which are discussed in the chapter on skill acquisition). The chapter on attention and control is relevant because we think that attentional focus affects our results, and the chapters on visual processing and motor control are relevant because not all games are played in a lab in front of a computer (we have done some testing on batting using professional baseball players to explore this). As discussed, the chapter on social simulations is relative because our approach is to treat games as interactions between agents. The chapter on personality and social psychology is also relevant because social situations can often be interpreted as games (we have applied our model to explaining childrens' friendship formation patterns) and, finally, the chapter on animal learning is relevant because competition is one of the primary ways that we understand animal behaviour and therefore we should consider if our model applies to animals. This multifaceted way of connecting is what Sloman referred to as "scaling out," and I think that most research in cognitive modeling can connect up in this way to many different areas. If we are to take the cognitive architecture paradigm seriously, this is something we will have to do much more of.

Since the chapters reviewing the different paradigms are important for framing the information in the rest of the book, I very briefly review them individually. Chapter 2 on connectionist networks, by Michael Thomas and James McClelland, provides an excellent overview and introduction to this area. It includes a discussion of the different types of connectionist networks and discussions of how connectionist networks relate to other systems, specifically, neural systems, Bayesian networks, and hybrid architectures. This chapter also includes three examples where connectionist networks made a big impact. A really good part of this section is that each contribution is followed by a discussion of how the model results changed how we think of the area. This is a sound illustration of how modeling is not just about matching a model to data. It is about gaining insight through the use of the model. So even if a model is eventually found not to be the whole story, it advances understanding and stimulates new avenues of research.

Chapter 3 on Bayesian models, by Thomas Griffiths, Charles Kemp, and Joshua Tenenbaum provides a clear explanation of this modeling paradigm. It can be read at a higher level, to get a conceptual understanding of what Bayesian modeling is, but it also has formulas that can be worked through to get a more detailed understanding. This chapter is important because Bayesian models are increasingly being used to model human cognition, and they are currently very important in the field of artificial intelligence.

Chapter 4 on dynamical systems, by Gregor Schöner, is, in my opinion, an excellent description of what dynamical computational models are about. Schöner also repeats the theme that modeling is about creating theoretical insights, not just finding the best way to fit the data. If you have read earlier books on this topic you will notice a big shift. Early on, in the 1990s, new ideas, such as embodiment, situated action, and dynamic systems theory were put forward as an alternative to computational models.

In particular, dynamic systems theory was put forward as a candidate to replace computation as the core modeling approach for understanding the mind (e.g., see van Gelder and Port, 1995). This has not come to pass because dynamic systems theory is not a distinctly different way of building models. In fact, any of the modeling systems mentioned in the book can be used to build dynamic models of cognition. However, as Schöner points out, the dynamic systems view has been instrumental in highlighting the importance of understanding how systems operate across time and within interactive environments. In addition, dynamic systems theory contains powerful theoretical tools for understanding and analysing this type of behaviour, which the reader can get a taste of by reading this chapter.

Chapter 5 on declarative/logic-based cognitive models, by Selmer Bringsjord, might be a bit confusing for the novice, as it attempts to do two things. As the name suggests, it examines attempts to model deliberate human reasoning processes (i.e., processes acting on declarative knowledge) using formal logic. The other issue is the relationship between the logic approach and cognitive architectures. Here, the point is that architectures tend to use rules as well as formal knowledge structures, such as propositions. Therefore, architectures can be represented as formal logic systems, with parameters added to deal with the fuzzier aspects of these systems. This may be true, but the people who use these systems do not understand them in this way. Bringsjord's argument is that having a formal way of representing the different modeling approaches (i.e., logic) would help the field, just as the common formalization of mathematics helps physics. Potentially, this could be a good thing, but only if this formalization led to clarification and insight, which remains to be seen. However, the point is well taken, we have difficulties talking to each other because we do not have a common language.

The last chapter in this section is "Constraints in Cognitive Architectures," by Niels Taatgen and John Anderson, and it has already been discussed. As noted above, it is not so much about describing specific architectures as it is about creating constraints to make them more testable. One connection that is interesting to notice though, is that the goal of establishing constraints based on perceptual/motor interactions with the environment is highlighted and discussed. This shows that the importance of understanding how human actions are situating in dynamic environments — as discussed in the chapter on dynamic systems — has been somewhat accepted by the mainstream modeling community.

Overall, I think this book is very helpful. Experimental psychologists tend to set up things so that their hypotheses can be tested within the span of a journal article. In contrast, for cognitive models, the span of a single journal article is only enough to show that the mechanism is plausible. To show real support for a mechanism requires many studies and comparisons to studies using different mechanisms. So it is very useful to have a book that looks across the various models and the findings that support them.

References

Anderson J.R., and Lebiere, C.L. (1998). *The atomic components of thought.* Mahwah, New Jersey: Lawrence Erlbaum.

Newell, A. (1990). *Unified theories of cognition.* Cambridge, Massachusetts: Harvard University Press.

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

Stewart, T.C., and West, R.L. (2006). Deconstructing ACT–R. *Proceedings of the Seventh International Conference on Cognitive Modeling* (pp. 298–303). Trieste, Italy.

Sun, R. (Ed). (2006). *Cognitive modeling and multi-agent interactions.* New York: Cambridge University Press.

van Gelder, T., and Port, R.F. (1995). It's about time: An overview of the dynamical approach to cognition. In R.F. Port and T. van Gelder (Eds.), *Mind as motion: Explorations in the dynamics of cognition* (pp. 1–44). Cambridge, Massachusetts: MIT Press.

West, R.L., Lebiere, C., and Bothell, D.J. (2006). Cognitive architectures, game playing, and human evolution. In R. Sun (Ed.), *Cognitive modeling and multi-agent interactions* (pp. 103–123). Cambridge: Cambridge University Press.