ARTICLE IN PRESS

Available online at www.sciencedirect.com

ScienceDirect

Cognitive Systems Research xxx (2008) xxx–xxx

Cognitive Systems Research

www.elsevier.com/locate/cogsys

Theoretical status of computational cognitive modeling

Action editor: Vasant Honavar

Ron Sun

Cognitive Sciences Department, Rensselaer Polytechnic Institute, Troy, NY 12180, USA

Received 20 April 2008; accepted 8 July 2008

Abstract

This article explores the view that computational models of cognition may constitute valid theories of cognition, often in the full sense of the term "theory". In this discussion, this article examines various (existent or possible) positions on this issue and argues in favor of the view above. It also connects this issue with a number of other relevant issues, such as the general relationship between theory and data, the validation of models, and the practical benefits of computational modeling. All the discussions point to the position that computational cognitive models can be true theories of cognition.

Keywords: Cognitive modeling; Cognitive architecture; Theory; Simulation; Validation

1. Introduction

With the increasing use of computational cognitive modeling and simulation in cognitive science, the theoretical and methodological status of computational cognitive modeling and simulation needs to be better understood. Despite some significant and intriguing computational cognitive modeling work in the past (see, e.g., Anderson & Lebiere, 1998; Meyer & Kieras, 1997; Sun, Merrill, & Peterson, 2001), larger methodological questions concerning such work remain. They are becoming more acute, given the steadily growing interest in computational cognitive modeling. For example, whether computational modeling/simulation is a viable means for providing scientific theories in general, and for cognitive science in particular, is an important question to consider. Furthermore, if it is a viable way of constructing and expressing scientific theories, how exactly is it to be used to produce scientific theories in cognitive science? What is its relationship with verbal–conceptual theories of cognition[1] or with mathematical theories of cognition? What is its relationship with data and observations that a scientific theory is supposed to explain (e.g., those produced by cognitive psychology or social psychology)? And so on (for similar or related issues, see also Durkheim, 1962; Kuhn, 1970; Newell, 1973; Poincare, 1982; Suppe, 1977; van Fraassen, 1980).[2]

Some believe that computational modeling/simulation does not explain anything per se, but (at most) confirms pre-existing general verbal–conceptual theories. Some further claim that simulations are built on top of pre-existing theories, and they may act as tests of those theories, but not in a truly significant way. This is because, they claim, simulation may not really validate a theory, but only perhaps help to falsify it sometimes, in case a narrowly-scoped the-

E-mail address: rsun@rpi.edu

[1] Examples of verbal–conceptual theories may include Gestalt psychology, Freudian psychology, Maslow's theory of human needs, or the drive theory of animal motivation.

[2] Here, "theory" is a rather fluid notion — it mostly involves an account of a set of phenomena (in some form) in terms of the productions and/or the regularities of these phenomena. On the other hand, the notion of "explanation" here involves the application of a theory to describe phenomena (so, it is derived from the notion of "theory").

1389-0417/$ - see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.cogsys.2008.07.002

ory and a correspondingly narrowly-scoped computational model are involved. Or this is because simulation may not help to falsify much, due to the generality, in case a broadly-scoped cognitive architecture is involved (e.g., Anderson & Lebiere, 1998). Moreover, if computational modeling/simulation is to have any bearing on a general verbal–conceptual theory (such as revealing inconsistencies, or providing empirical support), it must be logically derived from the theory. This requirement is believed by some to pose an almost insurmountable difficulty (as will be explicated later).

These issues deserve some further considerations. In this paper, I would like to discuss these issues, and a myriad of positions associated with them, in relation to the theoretical and methodological status of computational cognitive modeling and simulation. These issues are highly significant for the future development of computational cognitive modeling – which may lead up to a "computational psychology" (in the same way as some better-established computational fields such as computational biology). To benefit such future developments, it is highly desirable to establish that computational cognitive models are themselves cognitive theories.

In the remainder of this paper, first computational cognitive modeling in general is sketched. Next, the use of computational cognitive architectures in particular is touched upon in the context of computational cognitive modeling. Then, various positions concerning the nature of scientific theories are outlined. On the basis of such background, a number of positions concerning the theoretical and methodological status of computational cognitive modeling are explored in the subsequent section. Some unifying perspectives on different types of theories are then presented. From these perspectives, it is argued that the position that computational cognitive models may themselves be cognitive theories is well supported. Next, the practical value of computational cognitive modeling is accentuated, in terms of both precision and expressiveness, which further supports the position that computational cognitive models may themselves be cognitive theories.

2. Computational cognitive modeling

Before jumping into a discussion of the theoretical status, the explanatory utility, and the significance of computational cognitive modeling, a brief look at what it is and how it came about is in order.

Computational cognitive modeling provides detailed descriptions of mechanisms (i.e., static aspects) and processes (i.e., dynamic aspects) of cognition (Luce, 1995; Machamer, Darden, & Craver, 2000). It embodies descriptions of cognition in algorithms and programs, based on the science and technology of computing (Turing, 1950). That is, it produces runnable computational models. Detailed simulations can then be conducted based on the runnable computational models. Right from the official beginning of cognitive science around late 1970's, computational modeling has been a mainstay of cognitive science.

From Schank and Abelson (1977) to Minsky (1981), a variety of symbolic "cognitive" models were proposed in Artificial Intelligence. They were usually broad and capable of a significant amount of "cognitive" information processing. However, they were usually not rigorously evaluated against human data. Therefore, the cognitive validity of many of these models was not clearly established. Psychologists have also been proposing computational cognitive models, which were usually narrower in terms of scope and coverage. For instance, an early example was Anderson's HAM (Anderson, 1983). These models were usually more rigorously evaluated in relation to human data. Many of such models were inspired by symbolic AI work mentioned above (Newell & Simon, 1976).

The resurgence of neural network models in the early 80's brought another type of computational model into prominence in this field (e.g., Rumelhart et al., 1986). Instead of symbolic models that rely on a variety of complex data structures that store highly structured pieces of knowledge (such as Schank's Scripts or Minsky's Frames), simple, uniform, and often massively parallel numerical computation was used in these neural network models (Rumelhart et al., 1986). Many of these models were meant to be rigorous models of human cognition, and they were evaluated in relation to human data in a quantitative way (but see also Massaro, 1988).

Hybrid models that combine the strengths of neural networks and symbolic models emerged in the early 90's (see, e.g., Sun & Bookman, 1994; Wermter & Sun, 2000). Such models might be used to model a wider variety of cognitive phenomena due to their more diverse and more expressive representations (but see Regier, 2003). They have been used to tackle a wide range of cognitive data, often (though not always) in a rigorous way (see, e.g., Sun and Bookman, 1994; Sun, 2002).

Computational models can correspond to (i.e., match) actual human data in a variety of ways and can thereby be validated (at least theoretically). Computational cognitive models can be either broad or narrow (either covering a large set of data or being very specialized), precise or imprecise, and descriptive or normative. There are at least the following types of correspondences between computational models and human behaviors, in an increasing order of precision (Sun & Ling, 1998):

- Behavioral outcome modeling: A computational model produces roughly the same types of behaviors as humans do, under roughly the same circumstances. For example, given a set of scenarios for decision making, a model makes roughly the same broad kinds of decisions as a human decision maker.
- Qualitative modeling: A computational model produces the same qualitative behaviors that characterize human performance, under a variety of circumstances. For example, the performance of human subjects improves/

deteriorates when one or more control variables are changed; if a model shows the same changes given the same manipulations, one may say that the model captures the human data qualitatively.

- Quantitative modeling: A computational model produces exactly the same quantitative behaviors as exhibited by humans, as indicated by some quantitative performance measures. For example, one can perform point-by-point matching of the learning curve of a computational model and that of humans, or one can match step-by-step performance of a model with the corresponding performance of humans.

Exploring the match between a model and human data is an important means of understanding the human mind. Obtaining a good fit is not as trivial a result as one might believe. Finding a good fit often involves painstakingly detailed work. Moreover, validation of internal mechanisms and processes is even more difficult, which is a significant issue that is gaining increasing attention (e.g., Pew & Mavor, 1998; see more discussions later in Section 7).

Computational models have had some successes in terms of capturing and explaining a wide variety of cognitive phenomena in one of the above three senses. It is not an exaggeration to say that by now computational cognitive modeling in fact constitutes computational psychology as well as theoretical psychology (Newell, 1980, 1990; Sun, 2008). This is because developing a computational model and matching it against empirical human data is an important way of exploring human cognition. Finding a good fit involves detailed explorations of mechanisms and processes – the result is a detailed understanding of what affects performance in what ways (Thagard, 1986). It is the psychology that relies on computational model development as the essential methodological approach. Hence it constitutes a computational psychology (Sun, 2008). Computational modeling contributes to general, theoretical understanding of cognition through generating mechanistic and process-based descriptions that match human data (Luce, 1995). Therefore, it is also theoretical psychology – theorizing about cognition with (or through) computational means.

3. Cognitive architectures in computational cognitive modeling

One particularly important strand of computational cognitive modeling work is based on "cognitive architectures", that is, broadly-scoped, domain-generic computational cognitive models focusing on essential structures, mechanisms, and processes. They are used for broad, cross-domain analysis of cognition (Newell, 1990; Sun, 2002). A cognitive architecture provides a concrete framework for more detailed modeling of cognitive phenomena, through specifying essential structures, divisions of modules, relations among modules, and a variety of other essential aspects (Sun, 2002, 2004). They help to narrow down possibilities (i.e., the space of possible models) and to provide scaffolding (i.e., essential structures), through embodying fundamental theoretical assumptions (e.g., implicit versus explicit cognition, as in Sun (2002); see also Pew & Mavor, 1998; Ritter et al., 2003).

A computational cognitive architecture carries with it a set of structural (including ontological) assumptions, which form the basis of the architecture. On top of that, more detailed mechanisms and processes are specified for various cognitive faculties, such as memory of various forms, learning, and decision making. These mechanisms and processes are also the key elements of an architecture. These assumptions (structural, mechanistic, or process-based) represent a theoretical commitment, the implications of which can then be explored.[3] In addition, some parameters may be specified uniformly a priori in a cognitive architecture, which form a set of additional assumptions about cognition, on top of structural, mechanistic, and process-based assumptions. The basic structural assumptions and the specifications of essential mechanisms and processes may lead to the identifications of the values of many parameters, especially when these parameters have clear theoretical or empirical interpretations. Beside that, empirical data may provide a means for estimating values of other parameters (on the basis of prior theoretical assumptions, of course).

Looking back into the history of ideas concerning the architecture of the human mind, we see complementary lines of thoughts and gradual development of details, which led to the present-day computational cognitive architectures. For example, Kant viewed the human mind as a complex structure composed of innate faculties, and these innate faculties are then fine tuned by experience (as in, e.g., Anderson & Lebiere, 1998; Sun, 2002). Sigmund Freud focused attention on different subsystems of the mind and their interactions (as in, e.g., Sun, 2002, 2003). Allen Newell advocated the very idea of computational "cognitive architecture". One of the problems then was that both cognitive science and AI became fragmented, focusing mainly on specific issues, and lost sight of the big picture. Newell (1980, 1990) discussed a set of essential issues regarding architectures of the human mind, and argued that the field could make better progress if it addressed all of these issues together. This set of issues serves as a useful guide in that it puts emphasis on broad cognitive models that aim to capture general characteristics of cognition. Sun (2004) proposed another, more up-to-date and broader set of desiderata for developing generic cognitive architectures (see Sun, 2004 for details).

As a specific example, SOAR (Rosenbloom, Laird, & Newell, 1993) was a traditional symbolic representationalist cognitive architecture based on Newell's ideas. It attempted to capture an array of cognitive phenomena using a unified

[3] As in other forms of computational cognitive modeling, instead of purely mathematical equations, a combination of mathematical equations and computational procedures (algorithms) is usually used in cognitive architectures, with the advantages of being more flexible and more expressive. More on this later.

Please cite this article in press as: Sun, R., Theoretical status of computational cognitive modeling, *Cognitive Systems Research* (2008), doi:10.1016/j.cogsys.2008.07.002

mechanism – search through a state space, as well as associated explanation-based learning, based on elaborate symbolic representations (Newell, 1980, 1990).

On the other hand, ACT-R (Anderson, 1983; Anderson & Lebiere, 1998) has been the most influential and the most successful cognitive architecture, especially when compared with other symbolic representationalist models. ACT-R is made up of a semantic network (for declarative knowledge) and a production system (for procedural knowledge). Productions are formed through "proceduralization" of declarative knowledge, and have strengths associated with them, which are used for firing. Production firing is based on log odds of success. The essential point of ACT-R is that, in learning, declarative knowledge is acquired first, and then through practice, it is assimilated into procedural knowledge. The architecture involves elaborate symbolic (as well as numerical) representations.

Related in some ways to Anderson's ideas, some others (see, e.g., Albus, 1981; Sloman, 2000; Sun, 1999) divided cognition into several levels: reactive processing, deliberative processing, and reflective processing. In robotics, a similar division exists, which often includes three levels: the controller, the sequencer, and the deliberator. This division has been proposed by many roboticians over the years (see, e.g., Gat, 1998 for a review).

Along the same line, a more recent cognitive architecture, CLARION (Sun, 2002, 2003; Sun et al., 2001) employs two "levels" of representation, for capturing implicit and explicit knowledge, respectively. These two types of knowledge representations and their associated learning (of the two different kinds) enable complex and synergistic interaction between implicit and explicit cognitive processes. Learning is essentially "bottom-up", from implicit, reactive learning to the learning of elaborate symbolic conceptual structures, but it can go the other way around too ("top-down"). On top of these two kinds of representations, there are also meta-level cognitive processes and motivational processes (cf. Nelson, 1993).

The point of the foregoing discussion is that computational cognitive architectures make broad assumptions, and yet leave many details open. The essential considerations are often concerned with overall structural assumptions. Yet for practical reasons, a great deal of computational details (concerning mechanisms and processes) need to be specified. They need to be specified so that computational cognitive architectures can be compared to empirical data and be validated (among other things). The details are usually filled in during the course of the development of a cognitive architecture, through theoretical, computational, and/or experimental work.

4. Nature of scientific theories

Let us review some existing theories regarding scientific theories. First of all, according to scientific realism, product of scientific research is knowledge of mind-independent phenomena and such knowledge is possible even in those cases in which the relevant phenomena are not observable (Kitcher & Salmon, 1989). It entails that an ideal scientific theory makes true claims about genuinely existing unobservable entities. Realists believe that the operational success of a scientific theory lends credence to the idea that its unobservable aspects truly exist, as they were how the theory derived its predictions.

In contrast, constructive empiricism (van Fraassen, 1980) claims that scientific theories aim to be empirically adequate and their acceptance involves only the belief that they are empirically adequate. A theory is empirically adequate if everything that it says about observable entities is true, that is, consistent with empirical observations.[4] (Note that this is the view on which I shall base my arguments later.)

Against scientific realism, it has been argued that many previously successful scientific theories achieved their predictive success through the postulation of entities that later, even more successful theories showed did not exist. It thus may be argued that we should expect for theories to be replaced by newer ones that postulate a different set of unobservables (theoretical entities). Viewed in this light, scientific realism appears unjustified.

Relatedly, some have noted that data and observations are often (if not always) dependent on theories (that is, they are theory-laden; Kuhn, 1970). The fact that theories can only be tested as they relate to other theories implies that one can claim that test results that seem to refute a scientific theory have not refuted that theory at all. Rather, it is possible that the test results conflict with predictions because some other theory is false. Given this indeterminacy, scientific realism appears again unjustified.

Similarly, against scientific realism, social constructivists point out that scientific realism is unable to account for the rapid changes that occur in scientific knowledge during periods of scientific revolution (Kuhn, 1970). According to social constructivists, the success of scientific theories is part of the social construction of knowledge, as opposed to revealing some "objective" truth about reality as scientific realism would claim (Kukla, 2000). (I shall return to this issue of "construction" versus realism later in Section 7.)

The upshot, therefore, is that scientific realism should not be taken for granted, and may not be a solid philosophical foundation for science in general and for computational cognitive modeling in particular. At a minimum, we need to take into account its alternatives (such as constructive empiricism). Note that these issues are highly controversial and far from being settled. I am not going into

[4] According to van Fraassen (1980), "To present a theory is to specify a family of structures, its models; and secondly, to specify certain parts of those models (the empirical substructures) as candidates for the direct representation of observed phenomena. ...... the theory is empirically adequate if it has some model such that all [observed empirical] appearances are isomorphic to empirical substructures of that model" (p. 64).

Please cite this article in press as: Sun, R., Theoretical status of computational cognitive modeling, *Cognitive Systems Research* (2008), doi:10.1016/j.cogsys.2008.07.002

details of these issues, because they are not the focus of the present article. However, I do want to bring them up as the background of what I will be discussing.

Relatedly, there are also differing views on the interpretation of scientific theories. The linguistic view treats a scientific theory as a set of linguistic statements that are used to derive logical consequences (Hempel, 1965; Suppe, 1977). The semantic view, in contrast, treats a scientific theory as implying a general picture of the relevant phenomena – models of the relevant phenomena defined by all their relevant dimensions and by all their possible configurations across time (Suppe, 1977; van Fraassen, 1980). The models may be based on set theory, state spaces, or some other mathematical formalisms. It is important to note that the notion of "model" in the semantic view (e.g., Cartwright, 1997; Giere, 2004; Morgan & Morrison, 1999) is generally not the same as that used in computational cognitive modeling. The former is a generic notion that denotes some formal processes underlying interpretations of verbal–conceptual or mathematical scientific theories, while the latter is a specific notion denoting a type of elaborately developed formal descriptions (that may be coded as computer programs).[5] In other words, while the linguistic versus semantic distinction was concerned with the interpretation of scientific theories, my main concern here is instead the expression (the formalism) of scientific theories.[6]

Turning to another related issue, I would like to highlight Machamer et al. (2000)'s emphasis on the importance of "mechanisms" in scientific theorizing (see also Bechtel & Abrahamsen, 2005). According to Machamer et al. (2000), "in many fields of science what is taken to be a satisfactory explanation requires a description of a mechanism. So it is not surprising that much of the practice of science can be understood in terms of the discovery and description of mechanisms" (p. 1–2). "The contemporary mechanical world view, among other things, is a conviction about how phenomena are to be understood" (p. 21). To me at least, it implies that mechanisms can be a legitimate part of a scientific theory (or even a major part of a scientific theory), based on which mechanistic explanations of phenomena can be constructed.

In Machamer et al.'s use of the term, "mechanism" includes both entities as well as activities involving entities (i.e., including both static and dynamic aspects). Using the common terms of cognitive science, their notion of "mechanism" involves both representations as well as cognitive mechanisms and processes operating on them. Describing such cognitive "mechanisms" is evidently the main objective of cognitive science, and computational cognitive modeling in particular. So Machamer et al.'s view is fully compatible with the currently common and prevailing practices in cognitive science and in computational cognitive modeling.

In cognitive science, computational modeling is particularly suitable for describing such "mechanisms" of cognition (Thagard, 1986), with all of their details, in a constructive empiricist way (a la van Fraassen, 1980). Cognitive representations can be easily translated into and implemented in computer data structures, and cognitive mechanisms and processes can be implemented through computer algorithms. Any amount of detail of a "mechanism" in Machamer et al.'s sense (provided that it is Turing computable) can be described in an algorithm, while it may not be the case that it can be described through mathematical equations (that is to say, algorithms are more expressive). So, with the potentially huge amount of fine-grained details that can be involved in cognitive processes and mechanisms, computational modeling is especially important in revealing mechanistic and process details (albeit in a constructive empiricist way; more discussions on this point later in Section 7), and thus it is an important part of theoretical development in cognitive science. As warned by Machamer et al. (2000), "we should not be tempted to follow Hume and later logical empiricists into thinking that the intelligibility of activities (or mechanisms) is reducible to their regularity" (p. 21). "Rather, explanation involves revealing the productive relation" (p. 22). Similarly, Salmon (1998) pointed out that "What does explanation offer, over and above the inferential capacity of prediction and retrodiction....? It provides knowledge of the mechanisms of production and propagation of structure in the world. That goes some distance beyond mere recognition of regularities, and of the possibility of subsuming particular phenomena thereunder" (p. 29). See also Bechtel and Abrahamsen (2005). Detailed accounts of production are what is often provided by computational cognitive modeling. Furthermore, it may be argued that computational cognitive modeling revealing such "mechanisms" and "production" can be an important part of cognitive theories per se (as I shall argue next).

5. Theory versus model in cognitive science

With the afore-discussed background in mind, let us now consider the issue of the theoretical (and methodological) status of computational modeling and simulation: whether computational cognitive modeling constitutes theories of cognition, and whether something else is more essential for a theoretical understanding of cognition. Below I will identify and analyze a few common (or highly plausible) positions.

In relation to computational cognitive modeling, one possible (and starkly negative) viewpoint is that computational modeling and simulation, including those based on cognitive architectures, should not be taken as theory. A

[5] However, see the later discussion (in Section 5) of the view of computational models as instantiations of verbal–conceptual theories.

[6] It should be recognized that the boundary between theory and model is not clear-cut (even in the semantic view of scientific theories). Moreover, there are no clear-cut and commonly accepted definitions of theory and model that one can rely upon (Cartwright, 1997; Giere, 2004; Morgan & Morrison, 1999).

simulation is a generator of phenomena (i.e., generating various possibilities). Although it may be important for developing cognitive theories, it does not constitute cognitive theories, as some would claim. To produce a scientific theory of a given cognitive phenomenon, it is not enough to generate it. (Otherwise, as some would claim, any human would have a scientific understanding of the mind just because they produce cognitive phenomena.) To produce a scientific theory of cognition, it would be necessary to articulate it in a different (e.g., more "succinct") way.

Furthermore, as some would claim, a theory-building tool such as computational cognitive modeling would serve to build a theory, but it is not a theory. When looking for knowledge about a given phenomenon, one would prefer to study a theory than the tools used to build the theory. For one thing, if simulations and models are theories, they ultimately must conform to data; if they are only tools, then they need not (cf. Roberts & Pashler, 2000). If a model confirms the hypotheses obtained from interpreting a verbal–conceptual theory, then the model corroborates the theory; if it does not, it may lead to the formulation of new theories. In either case, the model may be a tool contributing to building a theory. Tools and theories should not be confused, as has been argued, because they serve different purposes. In this regard, it was claimed that computational cognitive models, cognitive architectures, modeling languages, and associated utilities were all tools.

A related claim is that computational modeling/simulation facilitates the precise instantiation of a (pre-existing) verbal–conceptual theory and consequently the careful evaluation of the theory against data. The scientific approach in the physical sciences is to demonstrate that a theory is empirically accurate or, wherever that demonstration cannot be made, that the theory is empirically adequate and those aspects that cannot be validated are useful (Bechtel, 1988; van Fraassen, 1980). As some would claim, computational modeling and simulation are useful in this sense (e.g., Axtell, Axelrod, & Cohen, 1996), but they are not theory per se because they are there for the sake of validating theories.

Along this line of models as tools, some would claim that they are tools for developing new theories: what is important is not models being derived from prior theories, but models leading up to new theories. The virtue of computational modeling and simulation (including cognitive architecture based modeling and simulation) is that the models can be rigorous, specific descriptors of observed behaviors, and when different observers of a behavior have different descriptions of that behavior, those different descriptions can all be modeled (cf. Sloman, 2000). This characteristic makes computational cognitive modeling and simulation different from traditional approaches. As some would claim, it enables us to engage and experiment with observations and data in a precise way, and possibly without the constraints of prior verbal–conceptual theories (which are often vague and not adequately validated). If good theories are to be produced, then the lesson from the physical sciences is that they need to be produced on the basis of good observations and data. As has been claimed, modeling enables us to formalize observations without spurious a priori generalizations. As such, it might be an important means of developing good cognitive theories. But still, it is not a theory itself because its form is not consistent with traditional forms of scientific theories and so on.

Yet another position is that different theories may be integrated into a simulation model. In constructing a computational simulation model, one often abstracts from the descriptive contents of verbal–conceptual theories and tries to formalize them into sets of equations or algorithms. In doing so, a computational simulation model may combine various theories (or various aspects of a theory). It therefore may lead to integrating different perspectives. That is, a computational model may specify when a subroutine with a particular set of equations or a particular algorithm is used (Sun, 2002, 2003; Sun et al., 2001). In this sense, the model "weighs" the perspective of each theory (or each aspect of a theory) in terms of its applicability and its relative impact. Thus, fragmentary theories compete and cooperate in the simulation model, and they also compete and cooperate in explaining simulation results. For example, ACT-R combines a theory of sensory-motor control with a theory of memory activation, among other things (Anderson & Lebiere, 1998). CLARION combines a theory of implicit and explicit learning with a theory of motivation, among other things (Sun, 2003). Further formulation of theories, in relation to simulation may lead to changing the theories, that is, improving the theories.

On the opposite end of the spectrum, there has been a radically different position, as advocated by, for example, Newell (1990) and Simon (1992), for which I shall argue. According to this position, a computational cognitive model can be a cognitive theory. A model can be a theory in and by itself. It is not the case that a model is limited to being built on top of an existing (verbal–conceptual or mathematical) theory, applied for the sake of generating data only, applied for the sake of validating an existing theory only, or applied for the sake of building a future theory only.

Let us see how we might arrive at such a position. Suppose one starts with a general verbal–conceptual theory (of, say, human motivation, or human decision making), and wants to construct a computational model and a simulation that are, in some sense, designed to reflect the essential explanatory structure of that theory. Invariably, one has to make numerous choices along the way, some of which are purely computationally motivated, in order to make the simulation run (especially with regard to which parameters are needed and which are not, and the functions that relate those parameters to one another in the computational model). Moreover, constructing a model often reveals logical gaps in the original theory that must be filled in order to make the simulation work. In the end, the model often introduces functional relationships that the original theory did not specify, and often specifies gap-filling assumptions that the original theory never made. These additions make the logical connection between the theory and the model tenuous at best. Computational models thus become distinct theoretical constructs. The fact that there may be some shared terms and some correspondences of functional relationships is not sufficient to establish a clear mapping between a theory and a computational model. Furthermore, there may not even be a consistent mapping. But, unless there is a clear and consistent mapping between a computational model and a theory, any "enhancement" that one adds to the verbal–conceptual theory is somewhat arbitrary. If there is no clear and consistent mapping between a computational model and a theory, then the output from the simulation has very little bearing on the original theory and cannot, in any rigorous sense, be a test of the original theory.[7]

More importantly, when constructing detailed simulation models inspired by "loose" conceptual–verbal theories, there are inevitably many "degrees of freedom" in specifying assumptions of the simulation models (usually much beyond minor indeterminacies such as yet-to-be-determined parameter values). If different people created different simulations based on the same theory, they would obtain rather different output patterns because of the ambiguity (major ambiguity beyond minor indeterminacies such as parameter values). The looser the original theory, the greater the chance that some of the simulations will contradict each other. We may be in a paradoxical situation that two modelers could apply the same theory differently, believing that their different simulation models (with different outcomes) were both valid representations of the theory.

If it is possible to have multiple, mutually contradictory simulation models derived from the same theory, then such a theory is underspecified, or even logically vacuous somehow, especially when it is underspecified to the extent that a significant portion of the output space is covered by different simulations derived from the same theory. The simulations in this case represent substantial enhancements over the original theory. If there are multiple simulation models with mutually contradictory results, then they are differently enhanced versions of the original theory, or different theories (at least from a logical standpoint). Thus, they may need to be adjudicated by empirical work. However, the falsification of any of these models may not truly falsify the original theory, because of the strenuous connections between them. (And in case a generic cognitive architecture is involved, falsification is even more difficult, because of the very generality of a cognitive architecture.)

This scenario leads directly to the belief that a computational simulation model may be, in fact, a separate theory. A computational cognitive model is a formal description of relevant cognitive phenomena. The language of a model is, by itself, a distinct symbol system for formulating the theory (Newell, 1990).[8] In particular, in substantiating a verbal–conceptual theory for the sake of simulation, any enhancement (of any of the two kinds discussed above) that one does to the verbal–conceptual theory is rather arbitrary. Instead, assuming that one has good definitions for the terms and the relationships involved in the simulation, it often makes more sense to think of the simulation model as a more rigorous theory, perhaps inspired by the original verbal–conceptual theory, but nonetheless a theory in its own right. No verbal–conceptual theory completely specifies the computational mechanisms involved, let alone the dynamic processes that may emerge. Thus, computational cognitive modeling is needed to specify these complex aspects, in order to produce a runnable computational simulation. The computational framework is, in essence, just another language for presenting a more rigorous and/or more detailed theory. Like verbal–conceptual theories or equation-based mathematical theories, computational models can, for example, be used to generate predictions. In fact, they can generate more precise predictions that can be more precisely (but not necessarily more easily) tested.

A logical conclusion from the above discussion is that a computational cognitive model can provide a cognitive theory and a scientific explanation of the corresponding cognitive phenomena. Their constructs and functional relationships can constitute the theory of, and the explanations for, the patterns of output they generate. This position has been advocated by many in the cognitive modeling community in the past (e.g., Anderson & Lebiere, 1998; Newell, 1990; Simon, 1992; Sun, 2006, 2008).

In addition to arguing against the "models as instantiations/derivations of verbal–conceptual theories" view, as has been done above, it is also necessary to argue against the other views mentioned above. For instance, in relation to the "models as data generators only" view, I want to emphasize that simulation models are often not just data generators, because there is rarely such a thing as a theory-free simulation (if ever). A computational simulation model often embodies theories, explicitly or implicitly (Kuhn, 1970), or it may constitute a theory by itself whether it is claimed to be so or not (Newell, 1990; Sun, 2008). Therefore, computational cognitive models and simulations need not be only a "data generator" (although they sometimes are) or a "theory-building tool" (although they can be). On the contrary, computational cognitive models and simulations often embody cognitive theories

[7] The problem with the view that simulations are tests of verbal–conceptual theories is that, often, there is no way other than argument to show that a simulation represents a theory.

[8] There is the process of assigning meanings to symbols and data structures involved in computational models. Just as a verbal–conceptual theory is useful insofar as the meanings of its terms are (more or less) shared by its users, so should the meanings of the terms of a computational model be shared by its users. To enable this sharing, definitions for any terms that might be understood in different ways by different researchers may be needed, the same with all types of theories.

Please cite this article in press as: Sun, R., Theoretical status of computational cognitive modeling, *Cognitive Systems Research* (2008), doi:10.1016/j.cogsys.2008.07.002

or are cognitive theories themselves, although, as theories, they may not be as "succinct" as other types of theories (more discussions of this point later in Section 6). Note that this is not to say that a computational model cannot be a data generator, but that it is often more than that.

In relation to the "models directly engaging data and leading up to new theories" view as mentioned earlier, I should emphasize that, as argued before, there is rarely such a thing as a theory-free observation (due to the theory-ladenness of observations; Kuhn, 1970), and furthermore, there is rarely such a thing as a theory-free model (Kuhn, 1970). Thus, there is rarely such a thing as directly describing data. Consequently, computational cognitive models often constitute new theories themselves, rather than leading up to new theories (although they may do so sometimes).

In reality, computational cognitive models are rarely entirely inspired by one generic verbal–conceptual theory. They are often constructed from integrating multiple sources: possibly including (often narrowly-scoped) equation-based mathematical theories from psychology, computational models from AI, philosophical ideas, and so on. It is especially likely that they integrate fragmentary (narrowly-scoped) theories and show how they may fit together into a computational model, which specifically addresses how components may interact with each other (see, e.g., Anderson & Lebiere, 1998; Sun, 2002). So, in relation to the "cooperation/competition of multiple theories" view mentioned earlier, one may argue that the resulting model is a theory that has likely lost its direct connections to the original theories that were combined into the model. The formulation of a computational cognitive model for the sake of simulation may lead to modifying the details of the constituting original theories and/or the details of their inter-relations and their interactions – that is, the computational model may constitute a brand new integrated theory (see, e.g., Sun, Slusarz, & Terry, 2005a).

Relatedly, in the field of Artificial Life, there have been similar debates about the status of simulations. The "physics model" view in artificial life is based on the idea that results from simulations can be used to generate new hypotheses. The "emergent thought experiment" view is based on the notion that simulation may be used as a complex thought experiment (i.e. one where the outcome of a condition can only be ascertained by a computer simulation), analysis of which helps to explore and inform modifications of a theory and then the thought experiment is conducted again, ad infinitum (Bedau, 1999; Di Paolo, Noble, & Bullock, 2000).

Likewise, in the field of Social Simulation, there have also been some similar debates about these points. One can easily find counterparts of the various positions enumerated above among social simulation researchers. For example, the position that a simulation is a theory per se, or the position that a simulation is only a test of a theory, and so on, can all be found in that field (see, e.g., Axelrod, 1984; Axtell et al., 1996; Castelfranchi, 2001; Johnson, 1989).

6. Unifying perspectives on scientific theories

Over the past several hundred years of modern science, the conception of scientific theory has been changing, as physics in particular has become more and more abstract (Salmon 1984, 1998). For example, a scientific theory (or an "explanation" based on a scientific theory) was commonly understood as identifying causes. However, the interest in a more sophisticated definition of scientific theory grew as many of the scientific theories of the twentieth century came to construe, construct, and depend upon theoretical constructs that are not directly observed by human senses (including, e.g., certain notions of mechanisms; Bechtel & Abrahamsen, 2005; Machamer et al., 2000).[9] One cannot legislate, based purely on past experience, what counts and does not count as a scientific theory (as long as it is within a reasonable range). When we speak of a scientific theory, we should only insist that what is invoked as a theory really be an account of whatever it is supposed to account for (Ross & Spurrett, 2004).

A unifying perspective on various forms of scientific theories (such as verbal–conceptual theories, mathematical theories, or computational models) can be centered on the notion of the descriptive complexity of a theory. A theory should represent our best knowledge regarding the nature of a class of phenomena. However, depending on domains, our best knowledge varies in terms of explanatory succinctness.

In some cases, a small and rigorous set of mathematical equations are able to express the structures and the regularities of a domain to a sufficient extent, approximating with an acceptable level of accuracy. For example, in physics, Newtonian classical mechanics is such a case. However, in some other cases, a succinct set of equations cannot be found that can express domain regularities and structures to a satisfactory extent. In that case, a more complex form of theory may be required. Computational models, in my view, are but a possible class of complex theories for such domains. Understanding the human mind is one domain in which, notably, no simpler form of theory is available or sufficient.

The question is: Are these two different classes of theories, mathematical equations and computational models, fundamentally different? After all, mathematical equations and computational models are both instances of the class of formal models. In that sense, they are not fundamentally different. But they are certainly different in some less fundamental ways. One difference is that mathematical models

[9] Note that "theory" or "explanation" are loose and elastic notions, almost as loose and elastic as the underlying notion of understanding (see, e.g., Keil, 2006). In my use of the terms, they simply mean a general account of a certain domain at a certain level of abstraction as well as the descriptions of relevant phenomena based on the general account (van Fraassen, 1980, chapter 5).

are easier to specify (in terms of length of description) while computational models often take longer descriptions to express. Another difference is that mathematical models are often in the closed-form (i.e., with the relationship between input and output variables apparent) while computational models in the open form. However, issues of matching, validation, and prediction are common to all formal models, whether mathematical or computational.[10]

Let us examine the issue of descriptive complexity, that is, the issue of description length, in some more detail. Kolmogorov complexity (Li & Vitanyi, 1997) is a mathematical measure of complexity, based on how many binary bits are needed on a theoretical model of computation (i.e., the Turing machine) to capture (i.e., to encode) a computational process (that is, to express an algorithm). More loosely, Kolmogorov complexity measures the minimum length of the description of an algorithm (or a program). See Li and Vitanyi (1997) for technical details. It is a solid, though often neglected, foundation upon which we may theorize about forms and types of scientific theories as well. For instance, in a way, much of the scientific enterprise may be viewed as aimed at creating descriptions (i.e., theories) of the phenomenological world at a lower and lower Kolmogorov complexity – leading to simpler and simpler theories (Bechtel, 1988). This is a major, if not ultimate, goal of science.

We need to answer the question of why we need to condense information to obtain shorter theories. This is a question that can be answered based on existing, well-known principles. We may draw upon the minimum description length principle – the principle that says that the simplest theory is the best theory, based on the conjecture that the simplest theory can maximally enhance generality or universality of descriptions (Li & Vitanyi, 1997). A succinct theory may reduce the number of functional relationships needed to explain the world: Some relationships may be merged and accounted for at a deeper level. The same can be said of the number of essential entities involved.

A key difference between different types of scientific theories, I believe, lies in the description length of a theory (and possibly, by extension, the numbers of individual entities and relationships required by the theory). In this context, often, the deeper the level of description, the shorter the description length of the theory. However, often, the deeper the level of description, the longer the description length of the explanation for a phenomenon (constructed out of the theory). That is, shorter description lengths of deeper theories lead to "denser" (longer, more complex) descriptions of the explanations of phenomena.

For example, as argued in Coward and Sun (2004), a description of the scientific explanation of cooking a meal at the chemical level would require the inclusion of hundreds or thousands of different chemicals interacting in different ways. At the atomic level, the description of the explanation of the same phenomenon would require the inclusion of many more atoms. The quantum mechanical level would require the inclusion of each of the individual electrons and nuclei, with each of these particles being described by a probability distribution. A description of the explanation of even the simplest macro-level phenomenon would be painstakingly long if fully constructed using deeper theories at a micro level. A description of the explanation of the same phenomenon at a deeper level should in general be expected to have a higher description length, or a higher Kolmogorov complexity.

In practice, however, full descriptions of explanations of high-level phenomena are almost never created at the more detailed (deeper) levels. For example, there has never been a description of the scientific explanation of cooking a meal constructed at the quantum mechanical level. Detailed descriptions of explanations are created for a number of limited but precisely defined phenomena (Coward & Sun, 2004). Humans can only handle a limited amount of (explicit) information at one time (Sun, 2002). They must use a higher-level theory for thinking about (explaining) broader phenomena, and then zero in on smaller areas in order to apply a deeper theory.

In contrast to this higher descriptive (Kolmogorov) complexity of phenomena, the description of a theory per se at a deeper level often has a lower Kolmogorov complexity (although not always). This often means that the number of different types of entities that are needed at a deeper level is often smaller than that at a higher level, and the number of different types of relationships at a deeper level is often smaller as well (Coward & Sun, 2004). Thus, theories at a deeper level may often be more "succinct". In other words, a theory per se at a deeper level may often have a shorter description length, or lower Kolmogorov complexity (although, in contrast, descriptions of explanations of phenomena constructed out of the theory may have higher Kolmogorov complexity).

The paradoxical effect of deeper levels, in part due to the contrast between the descriptive (Kolmogorov) complexity of theories and that of explanations constructed out of the theories, leads to compromises – Multiple levels of descriptions may be used and relationships among them are explored (Sun, Coward, & Zenzen, 2005b). For example, from social simulation models to models of individual cognition, and further to detailed physiological models of the brain, multiple levels co-exist and are explored simultaneously, for the sake of coping with complexity in the above two senses (see Sun et al., 2005b).

Viewed in this light, equation-based mathematical theories and computational theories are often (though not always) at the opposite ends of the spectrum. Equation-based theories per se often have shorter description lengths, but often generate longer explanations of actually observed cognitive phenomena (if such explanations are generated at

[10] Taber and Timpone (1996) and Sun (2008) expressed this view in their books, respectively. Both mathematical and computational models are included in Johnson (1989)'s book as well, although it is in a different (but related) field. See also Lave and March (1975).

all). Computational models often have longer description lengths, but may often in the end generate shorter explanations of actual cognitive phenomena (and often at a deeper level as well). Imagine the scenario of generating explanations for a complex skill learning task. A more succinct equation-based theory may have trouble dealing with the cognitive mechanisms and processes underlying it, while a more complex computational model may produce a relatively succinct description (simulation) of such processes. See, for example, Sun et al. (2001) for such a case. However, both types are legitimate scientific theories by virtue of the fact that they both provide explanations of phenomena at a certain level of abstraction (or, at multiple levels of abstraction in some cases; Sun et al., 2005b). See, for example, Sawyer (2003) and Parunak, Savit, and Riolo (1998) for similar discussions regarding computational models versus mathematical equations (though with regard to agent-based versus mathematical modeling in social simulation). On the other hand, verbal–conceptual theories may be useful at an initial stage of investigation, by ways of providing initial hypotheses. But they usually do not provide precise and detailed theories comparable to either computational models or equation-based mathematical theories. Therefore, verbal–conceptual theories cannot be compared to the other two types through the measure of descriptive (Kolmogorov) complexity.

In cognitive science, instead of purely mathematical equations trying to capture details of cognitive mechanisms and processes, more complex and more detailed computational models are often used, in order to provide more precise and more detailed descriptions. Current mathematics, developed to describe the physical universe, may not be sufficient for describing the complex human mind. Compared with scientific theories in other disciplines (for example, in physics), computational cognitive modeling may be mathematically less elegant (Greene, 1999). But the reality is that the human mind itself is likely to be less mathematically elegant compared with the physical universe, because the human mind is made up of complex artifacts that are created through a long and incremental evolutionary process (see Minsky, 1985 for a similar view). Therefore, an alternative form of theorizing is called for, a form that is more complex, more diverse, and more algorithmic (as opposed to mathematical) in nature. However, these two forms, mathematical and computational, can be complementary to each other. So they can co-exist in cognitive science, for example, for the sake of generating descriptions at different levels and of different scales. In fact, often, in cognitive science, a combination of mathematical equations and computational procedures (i.e., algorithms) is used, as currently commonly practiced in computational cognitive modeling. The advantages of using such hybrid forms are the added flexibility and the added expressive power such forms afford, which many researchers in the cognitive science community argue are necessary for understanding a system as complex as the human mind (Minsky, 1985; Newell, 1980, 1990; but see Massaro, 1988; Regier, 2003). Computational cognitive models provide a viable way (the only viable way, as some may argue) of specifying detailed, precise, and complex theories of cognition. Consequently, they may provide detailed interpretations and insights that no other experimental or theoretical approach can provide.

According to Kitcher (1981), establishment of common patterns in theoretical descriptions and explanations (‘‘argument patterns’’) that hold within (or across) sciences is necessary. For instance, evolutionary biology is characterized by its frequent use of explanations that cite measurable effects of environmental or other selection on the distribution of inheritable properties within populations. A biologist does not question the validity of this kind of theory in general, because accepting the soundness of its logic is part of what makes one a biologist. For computational cognitive modeling, what is currently lacking is the willing acceptance, within larger communities, of the approach as a completely valid form of theoretical description. (To some extent, there is also some lack of internal agreement on some details of acceptable theoretical-explanatory patterns within computational cognitive modeling.) But computational cognitive modeling can, and should, become accepted theoretical-explanatory patterns in cognitive science.

Algorithms, as well as symbols and data structures used therein, constitute a unique language of scientific theorizing, somewhat different from either verbal–conceptual theories or mathematical theories. The technology of computing apparently brought something new to scientific theorizing (Turing, 1950), in cognitive science, as well as in some other fields (such as social sciences). It appears that we may be currently in a period of ‘‘revolutionary’’ (i.e., transitional) science (as indicated by many before; Kuhn, 1970), in the sense of intensely building and heavily relying on new kinds of theorizing that traditional cognitive sciences (traditional experimental psychology, mathematical psychology, and so on) were not able to utilize (as a result of the lack of tools for algorithmic descriptions). Although other fields may be utilizing computational methods too, the significance of such methods in many of these fields (such as physics, chemistry, and so on) may not be as great as that of computational methods in cognitive science. In fact, computational modeling has been the very foundation of cognitive science, ever since its inception (Newell, 1980). Because of the adoption of this new kind of theorizing, it may in turn be possible to provide precise and detailed theories that traditional methods cannot.[11]

If we are in a transitional period dealing with new analytical tools (‘‘algorithms’’) and new intellectual approaches to theory building, it is necessary that we fully understand the power as well as the limitations of the new tools and approaches. It may be appropriate to delimit what we can do as well as how we can report results by a set of new criteria (see, e.g., Schunn & Wallach, 2001).

[11] There have been similar points made about computational modeling in some other related fields (such as the field of social simulation; see Moss, 1999).

Please cite this article in press as: Sun, R., Theoretical status of computational cognitive modeling, *Cognitive Systems Research* (2008), doi:10.1016/j.cogsys.2008.07.002

These criteria may emerge from the cumulative experience of the whole of the research community. New kinds of theorizing may not be bound by old ("traditional") criteria regarding scientific theories and, in particular, regarding forms of such theories.

A second (auxiliary) possibility for a unifying perspective on various forms of scientific theories is to keep in mind the trade-off between the descriptive (Kolmogorov) complexity of a theory and the scope of its empirical content (Lakatos, 1970). Given that at any particular level of abstraction, with increasing descriptive complexity of theories, more and more empirical data (observations) may often be accounted for, there is the question of when a theory (for example, a cognitive architecture) is too complex. This issue may be considered from the viewpoint of the "ratio" of complexity over empirical content (which is a sort of cost–benefit ratio). Obviously, we want to account for as much empirical content as possible. We also want to limit the growth of complexity of a theory as much as possible (or even try to reduce its complexity). Then, we need to trade-off the coverage of empirical content with the level of complexity. The difference between equation-based mathematical theory and computational modeling may be viewed, to some extent, in this light. Equation-based theories are generally more concerned with simplicity (i.e., reduction of complexity), while computational models (especially cognitive architectures) are more aimed for the breadth and scope of data/observations accounted for.

Yet another (auxiliary) possibility for a unifying perspective on various forms of scientific theories is also based on the complementarity of different methods. A complete scientific theory would address the what, when, how, and why of a phenomenon. When we have a phenomenon to explain, we would like to discover the necessary and sufficient conditions of that phenomenon. In the process, we may need to consider different degrees of category membership of empirical cases. We also need to investigate empirically the major entities and the major functional relationships involved. We may want to show how the mechanisms and processes involved work. In other words, we may come up with a detailed mechanistic and process-based explanation. Optionally, we may also need a deeper theory of why the phenomenon is present, and so on. So, a scientific theory requires at least: (1) a specification of the essential condition of a given phenomenon and (2) a specification of the mechanisms and processes by which the condition brings about the phenomenon. Computational cognitive modeling helps to delineate both of the above two aspects, but often more of the second aspect (see, e.g., Sun et al., 2001). On the other hand, equation-based cognitive theories and verbal–conceptual cognitive theories excel more at addressing the first aspect. Thus, different types of theories may be complementary to each other in this sense as well.

The discussions above present several unifying perspectives on different types of theories, and unequivocally support the view that computational cognitive models can be theories of cognition in and by themselves. Hopefully, these perspectives above will be a step towards resolving the dispute among the different positions (as enumerated earlier) on this issue.[12]

7. Relationship between Theories and Data

Let us turn to methodological issues. In connection with the afore-discussed unifying perspectives, we may examine claims about the significance of data and observations in theory formation. Some insist that one should produce, accept, or reject theories, strictly based on data and observations, and use this point in their arguments against computational cognitive modeling. However, there are plenty of reasons to believe that, without theories, observations and data are often impossible. There is rarely such a thing as a theory-free observation (Kuhn, 1970; van Fraassen, 1980). Without an explicitly stated theory, observations and data are laden with hidden, often inaccurate, vague, but intuitive theories (as often seen in experimental psychology). One does not start with a conceptually blank slate. Existing conceptual structures constrain how one perceives the world and formulates observations (Kuhn, 1970; van Fraassen, 1980).[13]

The over-emphasis of the significance of data and observations is not based on any well accepted account of the history of science or philosophy of science. As Jules Henri Poincare pointed out, science is built with data just as a house is built with bricks, but a collection of data cannot be called science any more than a pile of bricks can be called a house. Theories do not simply accumulate from data and observations; instead, theories help us to select, organize, and integrate data and observations (Poincare, 1982).

For example, Einstein was, for the most part, not motivated by experimental data. Bodanis (2000) stressed the idea that the mathematical theory that Einstein proposed was not driven by data in any real sense. "What they [other physicists] could not grasp was that he [Einstein] didn't have any labs. The 'latest findings' he worked with came from scientists who'd died decades or even centuries before. But that did not matter. Einstein hadn't come up with his ideas by patiently putting together a range of new results. Instead, as he saw, he just spent a long time dreamily thinking about light and speed and what was logically possible in our universe and what wasn't" (p. 80).

Likewise, Kepler mostly did not build up the theory of the elliptical orbit from data. The people working from experimental data were not making progress on theories.

[12] Note that here I am not at all concerned with meta-physical issues. Thus, for example, I am not concerned with how computational theories may or may not be reduced to physics; how computational theories ultimately may or may not account for consciousness; and so on (cf. Coward & Sun 2004; Ross & Spurrett 2004). Overall, I am taking a pragmatic approach in this regard.

[13] By the same token, working with domain experts in building models constrains the models to conform to the experts' perception (to some degrees), which does not make the models accurate descriptions of objective reality, only (maybe) of the "expert" perception.

Only by thinking in a different way, theoretically, was progress made. Before Kepler, circular motion was essential to the concept of planet, and non-circular planetary orbits were simply inconceivable.

Copernicus's model of the solar system was also mostly theory-driven, not data-driven. At the beginning, the predictions of his model were not even as accurate as the earth-centered view. Newton also did not deduce theories from data, but proposed theories aimed to be compatible with empirical observations (van Fraassen, 1980). Kuhn (1970) contained ample evidence of such cases.

Most leading scholars in philosophy of science rejected the idea that new theories come about by building observations and data into patterns (Bechtel, 1988; Thagard, 1986; Toulmin, 1960; van Fraassen, 1980). Rather, the opposite is true: Theory is necessary in order to understand and organize data and observations. In physics, it is of no use even beginning to look at things until one knows exactly what one is looking for: Observation has to be strictly controlled by reference to some particular theoretical problem (Toulmin, 1960). The emphasis on using observations and data alone (or for the most part) to build new theories is misplaced.

The methodological implication of the above for computational cognitive modeling is that, in the course of computational cognitive modeling, proposing mechanisms and processes not strictly derived from data and observations (such as certain structural, mechanistic, or process-based assumptions in cognitive architectures as discussed earlier in Section 3) is justifiable methodologically. Remember that theories are needed in gathering data/observations and in interpreting them, so theories not strictly derived from data/observations are there to begin with. Detailed process/mechanism specifications in computational cognitive modeling can be viewed as theoretico-computational postulates, in a way equivalent to Kepler's or Copernicus' initial theories, which may then be evaluated in terms of empirical adequacy (van Fraassen, 1980) on the basis of data/observations that they are aimed to account for. These data/observations may be collected under the guidance of, in correspondence to, and sometimes for the sake of empirical validation of the postulates embodied in the process/mechanism specifications of computational cognitive modeling. As pointed out by van Fraassen (1980), "empirical minimality is emphatically not to be advocated as a virtue, it seems to me. The reasons for this point are pragmatic. Theories with some degree of sophistication always carry some 'metaphysical baggage'. Sophistication lies in the introduction of detours via theoretical variables to arrive at useful, adequate, manageable descriptions of the phenomena" (p. 68). Judging from the physical sciences, such sophistication is necessary to arrive at deep and broad scientific theories. The fact that some details of such theories were motivated from a computational standpoint need not be a problem, considering all the discussions above, because computational considerations are but one class of theoretical considerations (see also Thagard, 1986).

This implies that there are good reasons for using computational cognitive modeling, as well as that there are severe constraints on how computational cognitive modeling may proceed (just like other types of scientific theorizing). That is,

- Building computational cognitive models that contain speculative computational elements is not just inevitable, but in fact methodologically justified.
- Computational cognitive models and simulations of cognitive data are always theory-laden. When we make a claim based on a computational model or a simulation, we are taking a theoretical position, not simply observing the world.
- Assessing the validity of a computational cognitive model, in the scientific realist sense, is impossible. Instead, we can only aim for practical (i.e., empirical) adequacy (in the constructive empiricist sense); that is, a computational cognitive model is assumed to be valid until it is discovered to be no longer useful (e.g., replaced by a more accurate model).

Accordingly, building theories and models – computational cognitive models in particular – strictly from data and observations runs into the twin problems of the implausibility that one can have theory-free data and observations and, on top of that, the difficulty of establishing that the theories and models are valid representations of such data and observations.

It should be emphasized that constructive empiricism, the philosophical view that scientists customarily build theories that are compatible with empirical data and observations rather than strictly engage in deducing underlying "reality" from data and observations, may make a more sensible philosophical foundation for computational cognitive modeling, especially when it is compared with scientific realist accounts (believing that scientific theories necessarily reflect the "true reality") or naive empiricist accounts (limiting theories to be strictly based on empirical data and observations) that some scientists seem to subscribe to. Like other scientific theories, computational cognitive models can be viewed, in some sense, as plausible and useful fictions (with regard to the unobservables; van Fraassen, 1980). See van Fraassen (1980) for a detailed account of this position (see also the discussion in Section 4 earlier).

This issue of theory versus data/observation may be related to the Kuhnian notion of "paradigm", within the unifying perspectives outlined earlier. According to Kuhn (1970), on the assumption that a current theory is consistent and correct, observations and data are collected and fitted within this current theory (or scientific "paradigm"). New and unexpected phenomena may be uncovered in the process, and they may lead to revision and refinement of the existing theory. In cognitive modeling, especially with cognitive architectures, bold initial hypotheses (e.g., regarding a cognitive architecture) are made, and gradual refinement and re-organization of details follow. That is, structural and other commitments (such as mechanistic and process-based assumptions in a cognitive architecture) constitute an initial theory, which undergoes testing and validation through matching with data. Revision and refinement are undertaken when inconsistencies with, or incorrect predictions by, a model are discovered, or when the model is incapable of predicting something important (Lakatos, 1970). However, when given a sufficiently high degree of mismatch between data and the current cognitive model, that is, when revision and refinement appear no longer able to accommodate problems that arise, a crisis may develop, which leads to a new "paradigm", for example, a new cognitive architecture or even a new approach towards building cognitive architectures. This process does not rely on the questionable existence of pure, theory-free data and observations, but views computational cognitive models as "paradigms", or provisional guiding principles, for experimental design, data gathering, and data interpretation, which themselves are subject to revisions in the process of directing experimental design, helping data gathering, and interpreting experimental results.

This discussion immediately brings up the practical issue of actual empirical validation of models (that is, in a constructive empiricist sense). Despite the argument made earlier regarding the inevitability of proposing mechanisms/processes that are not strictly empirically derived, some kind of empirical validation, in the sense of testing empirical adequacy (van Fraassen, 1980), after mechanisms and processes (such as structural, mechanistic, or process-based assumptions in cognitive architectures) have been proposed, is indeed needed in order to make computational cognitive modeling a science. In addition, in practical terms, in order for it to be convincing to the larger communities of scientists, it needs validation.[14]

A variety of practical methodologies for validating computational cognitive models have been developed, including using direct behavioral measures (such as accuracy, response time, and so on, mostly for validating input–output behaviors) and using indirect measures (such as eye tracking, brain imaging, and so on, mostly for validating internal details). However, computational cognitive models are notoriously hard to validate, even more so than other types of theories of cognition, because computational cognitive models specify much more details, especially computational details, of cognitive processes and mechanisms. The level of details involved makes them very hard to be validated in a systematic and methodical way (let alone completely). It appears that, for the sake of validation (and for other purposes), complexity reduction is an important issue (Bechtel & Richardson, 1993). As this is an important and broad topic, it deserves a separate treatment by itself; see Sun (2006a) for some details.

[14] However, note that I said "some kind of empirical validation", not "complete" validation, which is, as argued before, theoretically naive and practically infeasible (van Fraassen 1980).

Furthermore, in relation to validation, we might, at a theoretical level, draw upon ideas from Kuhn (1970), Lakatos (1970), and Laudan (1979), as mentioned before, in countering some criticisms that I would consider unjustified. A very general idea from these authors is that not every aspect of a scientific theory is, or can be, validated. A theory, as a whole, may be taken as a "paradigm", until it has been shown to be grossly inadequate (Kuhn, 1970), notably not just for lacking "complete" validation. Lakatos (1970) delineated the condition under which a theory (or a research program) may be considered inadequate, again not requiring complete validation of the theory (or the research program). The upshot is that the criticisms of computational cognitive modeling (including cognitive architectures) on the basis of involving a great deal of computationally motivated details and the difficulty with "complete" validations of these details are rather misplaced, and unjustified from a broad perspective derived from the history of science.

8. Practical benefits of computational cognitive modeling

Given the (mostly theoretical) discussions above, let us now turn to the practical importance of computational cognitive modeling in understanding cognition. I would like to enumerate a few important practical benefits of computational cognitive modeling. Admittedly, the existence of practical benefits does not directly justify treating computational cognitive models as theories, but it may make the view more palatable in a practical sense.

There are practical reasons to believe that the goal of understanding the human mind strictly from observations of human behavior is ultimately untenable (except perhaps for understanding human performance in small and limited task domains). The rise and fall of behaviorism is a case in point. The key point is that the mechanisms and processes of the mind cannot be understood purely on the basis of behavioral experiments, with tests that inevitably amount to probing only relatively superficial features of human behavior, which are further obscured by individual/group differences and contextual factors (Newell, 1973). It would be extremely hard to understand the human mind in this way, just like it would be extremely hard to understand a complex computer system purely on the basis of testing its behavior, when we do not have any a priori ideas about the nature, the inner working, and the theoretical underpinnings of that system. Therefore, theoretical developments (of various sorts) need to go hand-in-hand with experimentation on human behavior, as argued (in a more theoretical manner) in the previous section. This case may also be argued on the basis of analogy with the physical sciences; see Sun et al. (2005b) for details of such an argument.

Given the complexity of the human mind, and its manifestation in behavioral flexibility, complex mechanistic, process-based theories, that is, computational models, are necessary to explicate the intricate details of the mind. Without such complex theories, experimentation may be blind – possibly leading to the accumulation of a vast amount of data without any apparent purpose or any apparent hope of arriving at a concise and meaningful understanding. This is a serious pitfall, and in fact descriptive of some areas of empirical cognitive research (although not all).

It is true that even pure experimentalists may often be guided by their intuitive theories in designing experiments and in generating hypotheses. So, they are not completely blind. They may even be guided by specifically developed verbal–conceptual theories (see, e.g., Reber 1989 for an instance of such theories). However, in general, verbal–conceptual theories are often vague, and only intuitively suggestive – hence there lies another serious pitfall. Without detailed computational models, most of the details of an intuitive or a verbal–conceptual theory are left out of consideration. Nevertheless, there are many reasons to believe that the key to understanding cognition is often in fine details (see, for example, Sun et al., 2001, 2005), which one may argue only computational modeling can help to bring out. As pointed out by Hintzman (1990), "the common strategy of trying to reason backward from behavior to underlying processes (analysis) has drawbacks that become painfully apparent to those who work with simulation models (synthesis). To have one's hunches about how a simple combination of processes will behave repeatedly dashed by one's own computer program is a humbling experience that no experimental psychologist should miss" (p. 111). Computational models provide algorithmic specificity: detailed, precisely specified, and carefully thought-out steps, arranged in precise and yet flexible sequences. Therefore they provide conceptual clarity and precision at the same time. Without such detailed theories, empirical work often suffers from the two pitfalls mentioned above (even though vague forms of theories often do exist).

In this regard, let us look into some more details. There are several fundamental problems associated with verbal–conceptual cognitive theories. The first is the problem of the inevitability of omissions and inconsistencies. As human beings, cognitive scientists may fail to think of some factors and circumstances in complex situations. Thus, computational modeling and simulation may be needed to analyze the details of a process or a mechanism, in order to achieve a more thorough understanding (e.g., Sun et al., 2005a). The second problem is that of conflicting processes (e.g., Meyer & Kieras, 1997). There are different processes involved to different degrees in generating one phenomenon. A systematic and controlled analysis of all the involved processes is the basis of understanding a phenomenon. Computational modeling and simulation may be uniquely suitable for this purpose and thus indispensable in this regard (as discussed earlier, towards the end of Section 5). The third problem is that of variability of results. Some cognitive mechanisms may behave quite differently in different situations or with different parameter values. Thus, computational modeling and simulations may be needed to analyze the relevant parameters of a generic mechanism in a controlled manner, or to test it in a variety of different situations (e.g., Sun & Naveh, 2004).

Another useful characteristic of computational modeling and simulation is that they often produce not just one aspect of a situation but many different aspects. For example, in a cognitive model of skill learning, we may be dealing with accuracy of performance, response time, eye movement, hand movement, memory retrieval, and so on, all in one simulation model (e.g., Anderson & Lebiere, 1998; Sun, 2002). This comprehensiveness of computational cognitive modeling (especially when using cognitive architectures), as opposed to the usual specificity of mathematical theories, may be important for developing broad cognitive theories.

Also, computational cognitive models are useful media for thought experimentation. We may use simulations for exploring various possibilities regarding details of a cognitive phenomenon. As pointed out by Hintzman (1990), "a simple working system that displays some properties of human memory may suggest other properties that no one ever thought of testing for, may offer novel explanations for known phenomena, and may provide insight into which modifications that the next generation of models should include" (p. 111). (Mathematical models may be more limited and less useful in these regards, because of their usual simplicity.)

In particular, computational models allow the running of simulations for the sake of discerning consequences. Without the level of details as specified in computational models, it may be extremely difficult to see the consequences and implications of certain aspects of a complex model (Di Paolo et al., 2000). Sun and Zhang (2004) represents such an exploration. In the work, a number of simulations were conducted with variations of a computational cognitive model. Through systematic variations of a number of aspects of the model, conclusions were drawn regarding what an appropriate model for the particular task domain under consideration should be like with regard to these aspects considered. In general, differing details of different computational theories/explanations can be tested and compared through running computational simulations and then examining the results from the simulations. Even in the case when there is a mathematical model available, computational modeling and simulation can still be useful for assessing the consequences of various assumptions. Computational modeling and simulation can be used to ascertain whether or not an assumption is a key assumption. Beyond assessing the consequences of individual assumptions, computational modeling and simulation are appropriate for assessing the interactions of multiple assumptions. Since many cognitive theories (especially cognitive architectures) are based on multiple (often independently made) assumptions, the empirical validation of them, especially in terms of their interactions, is important, and can be readily addressed through computational modeling and simulation (Sun & Zhang, 2004; Sun, Zhang, Slusarz, & Mathews, 2007). (This characteristic of compu-

Please cite this article in press as: Sun, R., Theoretical status of computational cognitive modeling, *Cognitive Systems Research* (2008), doi:10.1016/j.cogsys.2008.07.002

tational cognitive modeling is relevant to addressing the three fundamental shortcomings of verbal–conceptual theories mentioned earlier.)

Even when a set of mathematical equations can be solved analytically, there may still be use for computational models and simulations. When closed-form analytical solutions exist, it appears that there is no need for simulation because the solutions have been completely specified. However, the dynamics implied by the equations may not be apparent. Moreover, many people do not work well with mathematical equations, but most people are adept at understanding concrete outcomes produced by a simulation (Keil, 2006). Especially when visual graphics is produced, it is easier for people to understand simulations, because it is often easier to perform visual pattern matching and recognition than abstract analysis (see Thagard, 2005 regarding the importance of visual thinking in science). This aspect of modeling/simulation is important particularly because cognitive science is an interdisciplinary field and made up of researchers with diverse background. Thus, there is certainly room for using computational modeling/simulation for communicative and pedagogical purposes.

In sum, the value of computational cognitive modeling (including using cognitive architectures) can be argued in many different ways, including in practical ways. See also, for example, Newell (1990), Sun (2002), Anderson and Lebiere (1998), Sun (2006, 2008), and so on. Computational cognitive models are more than just simulation tools, or programing languages of some sort (Newell, 1990). They are theoretically pertinent, because they may express theories in a unique and, I believe, indispensable way. Cognitive architectures, for example, may constitute broad, generic theories of cognition (as recognized by many in the field).

9. Concluding remarks

One conclusion that may be drawn from the foregoing discussions is that the theoretical and methodological status of computational cognitive modeling is undoubtedly significant. Computational cognitive modeling may provide not just tools, data generators, instantiations, or integrations, but also theories in the full sense of the term. It often provides detailed, mechanistic, process-based theories that enable the exploration of fine-grained details of cognitive phenomena, and serves as a useful guide to experimentalists who explore fine-grained details of cognitive phenomena through experimental means. In addition, it can be useful in a variety of other ways, theoretical or practical (Sun, 2008; Sun et al., 2005b). Computational models can be compared to both mathematical theories and verbal–conceptual theories, and excel as theories in terms of expressiveness and precision.

Acknowledgments

This work was carried out while the author was supported in part by Army Research Institute Contracts DASW01-00-K-0012 and W74V8H-05-K-0002 (to Ron Sun and Bob Mathews) and ONR Grant N00014-08-1-0068 (to Ron Sun). The author benefited from discussions with L. Andrew Coward, Michael Zenzen, and others. Paul Thagard provided valuable comments.

References

Albus, J. (1981). Brains, behavior, and robotics. Byte books. Peterborough, NH: McGraw Hill.

Anderson, J. R. (1983). The architecture of cognition. Cambridge, MA: Harvard University Press.

Anderson, J., & Lebiere, C. (1998). The atomic components of thought. Mahwah, NJ: Lawrence Erlbaum Associates.

Axelrod, R. (1984). The evolution of cooperation. New York: Basic Books.

Axtell, R., Axelrod, J., & Cohen, M. (1996). Aligning simulation models: A case study and results. *Computational and Mathematical Organization Theory, 1*(2), 123–141.

Bedau, M. (1999). Can unrealistic computer models illuminate theoretical biology? In A. Wu (Ed.), *Proceedings of 1999 genetic and evolutionary computation conference workshop program* (pp. 20–23). San Francisco: Morgan Kaufmann.

Bodanis, D. (2000). $E = mc^2$: A biography of the world's most famous equation. New York: Berkley Books.

Bechtel, W. (1988). Philosophy of science: An overview for cognitive science. Hillsdale, NJ: Erlbaum.

Bechtel, W., & Richardson, R. C. (1993). Discovering complexity: Decomposition and localization as strategies in scientific research. Princeton: Princeton University Press.

Bechtel, W., & Abrahamsen, A. (2005). Explanation: A mechanistic alternative. *Studies in History and Philosophy of Biology and Biomedical Sciences, 36*, 421–441.

Cartwright, N. (1997). Models: The blueprints for laws. *Philosophy of Science, 64*, 292–303.

Castelfranchi, C. (2001). The theory of social functions: Challenges for computational social science and multi-agent learning. In Ron Sun (Ed.), *Cognitive systems research, special issue on multi-disciplinary studies of multi-agent learning* (Vol. 2(1), pp. 5–38).

Coward, L. A., & Sun, R. (2004). Criteria for an effective theory of consciousness and some preliminary attempts. *Consciousness and Cognition, 13*, 268–301.

Di Paolo, E., Noble, J., & Bullock, S. (2000). Simulation models as opaque thought experiments. In *Artificial life VII: The 7th international conference on the simulation and synthesis of living systems*. (pp. 497–506). Cambridge, MA: MIT Press.

Durkheim, W. (1895/1962). The rules of the sociological method. Glencoe, IL: The Free Press.

Gat, E. (1998). On three-layered architecture. In D. Kortenkamp, R. Bonasso, & R. Murphy (Eds.), *Artificial intelligence and mobile robots*. Menlo Park, Calif: AAAI Press.

Giere, R. (2004). How models are used to represent reality. *Philosophy of Science, 71*, 742–752.

Greene, G. (1999). The elegant universe. Norton.

Hempel, C. (1965). Aspects of scientific explanation. New York: The Free Press.

Hintzman, D. (1990). Human learning and memory. *Annual Review of Psychology, 41*, 109–139.

Johnson, P. E. (Ed.) (1989). *Formal theories of politics: Mathematical modeling in political science*. Oxford, UK: Pergamon Press.

Keil, F. (2006). Explanation and understanding. *Annual Review of Psychology, 57*, 227–254.

Kitcher, P. (1981). Explanatory unification. *Philosophy of Science, 48*, 507–531.

Kitcher, P., & Salmon, W. (Eds.) (1989). *Scientific explanation*. Minneapolis, MN: University of Minnesota Press.

Please cite this article in press as: Sun, R., Theoretical status of computational cognitive modeling, *Cognitive Systems Research* (2008), doi:10.1016/j.cogsys.2008.07.002

Kuhn, T. (1970). Structure of scientific revolutions. Chicago: University of Chicago Press.

Kukla, A. (2000). Social constructivism and the knowledge of science. London: Routledge.

Lakatos, I. (1970). Falsification and methodology of research programs. In I. Lakatos & A. Musgrave (Eds.), *Criticism and the growth of knowledge*. Cambridge: Cambridge University Press.

Laudan, L. (1979). Historical methodologies: An overview and manifesto. In P. Asquith & H. Kyburg (Eds.), *Current research in the philosophy of science* (pp. 40–54). Philosophy of Science Association.

Lave, C. A., & March, J. G. (1975). An introduction to models in the social sciences. New York: Harper and Row.

Li, M., & Vitanyi, P. (1997). An introduction to Kolmogorov complexity and its applications. Heidelberg, Germany: Springer.

Luce, R. D. (1995). Four tensions concerning mathematical modeling in psychology. *Annual Review of Psychology, 46*, 1–26.

Machamer, P., Darden, L., & Craver, C. (2000). Thinking about mechanisms. *Philosophy of Science, 67*, 1–25.

Massaro, D. (1988). Some criticisms of connectionist models of human performance. *Journal of Memory and Language, 27*, 213–234.

Meyer, D., & Kieras, D. (1997). A computational theory of executive cognitive processes and human multiple-task performance: Part 1, basic mechanisms. *Psychological Review, 104*(1), 3–65.

Minsky, M. (1981). A framework for representing knowledge. In J. Haugeland (Ed.), *Mind design* (pp. 95–128). Cambridge, MA: MIT Press.

Minsky, M. (1985). The society of mind. New York: Simon and Schuster.

Morgan, M., & Morrison, M. (Eds.) (1999). *Models as mediators: Perspectives on natural and social science*. New York: Cambridge University Press.

Moss, S. (1999). *Relevance, realism and rigour: A third way for social and economic research. CPM Report No. 99-56*. Manchester, UK: Center for Policy Analysis, Manchester Metropolitan University.

Nelson, T. (Ed.) (1993). *Metacognition: Core readings*. Allyn and Bacon.

Newell, A. (1973). You can't play 20 questions with nature and win: Projective comments on the papers of this symposium. In W. Chase (Ed.), *Visual information processing* (pp. 283–308). New York: Academic Press.

Newell, A. (1980). Physical symbol systems. *Cognitive Science, 4*, 66–83.

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

Newell, A., & Simon, H. (1976). Computer science as empirical inquiry: Symbols and search. *Communication of ACM, 19*, 113–126.

Parunak, H., Savit, R., & Riolo, R. (1998). Agent-based modeling vs. equation-based modeling: A case study and user's guide. In J. Sichman, R. Conte, & N. Gilbert (Eds.), *Multi-agent systems and agent-based simulation*. Berlin: Springer.

Pew, R., & Mavor, A. S. (Eds.) (1998). *Modeling human and organizational behavior: Application to military simulations*. Washington, DC: National Academy Press.

Poincare, J. H. (1982). The foundations of science. Washington, DC: University Press of America.

Reber, A. (1989). Implicit learning and tacit knowledge. *Journal of Experimental Psychology: General, 118*(3), 219–235.

Regier, T. (2003). Constraining computational models of cognition. In L. Nadel (Ed.), *Encyclopedia of cognitive science* (pp. 611–615). London: Macmillan.

Ritter, F., Shadbolt, N., Elliman, D., Young, R., Gobet, F., & Baxter, G. (2003). *Techniques for modeling human performance in synthetic environments: A supplementary review*. Dayton, OH: Human Systems Information Analysis Center, Wright-Patterson Air Force Base.

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

Rosenbloom, P., Laird, J., & Newell, A. (1993). *The SOAR papers: Research on integrated intelligence*. Cambridge, MA: MIT Press.

Ross, D., & Spurrett, D. (2004). What to say to a skeptical metaphysician: A defense manual for cognitive and behavioral scientists. *Behavioral and Brain Sciences, 27*(5).

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

Salmon, W. (1984). *Scientific explanation and the causal structure of the world*. Princeton: Princeton University Press.

Salmon, W. (1998). *Causality and explanation*. New York: Oxford University.

Sawyer, R. (2003). Multiagent systems and the micro-macro link in sociological theory. *Sociological Methods and Research, 31*(3), 325–363.

Schank, R., & Abelson, R. (1977). *Scripts, plans, goals, and understanding: An inquiry into human knowledge structures*. Hillsdale NJ: Erlbaum.

Schunn, C., & Wallach, D. (2001). In defense of goodness-of-fit in comparison of models to data. *Manuscript*.

Simon, H. (1992). What is an "explanation" of behavior? *Psychological Science, 3*(3), 150–161.

Sloman, A. (2000). Architectural requirements for human-like agents both natural and artificial. In K. Dautenhahn (Ed.), *Human cognition and social agent technology*. Amsterdam: John Benjamins.

Sun, R. (1999). Accounting for the computational basis of consciousness: A connectionist approach. *Consciousness and Cognition, 8*, 529–565.

Sun, R. (2002). *Duality of the mind*. Mahwah, NJ: Lawrence Erlbaum Associates.

Sun, R. (2003). A tutorial on CLARION. Technical report, Rensselaer Polytechnic Institute. <http://www.cogsci.rpi.edu/~rsun/sun.tutorial.pdf>.

Sun, R. (2004). Desiderata for cognitive architectures. *Philosophical Psychology, 17*(3), 341–373.

Sun, R. (Ed.) (2006). *Cognition and multi-agent interaction*. New York: Cambridge University Press.

Sun, R., (2006a). What is validation of cognitive architectures and is it possible? Paper presented at the Air Force Workshop on model comparison and model validation. Syracuse, NY, September 8–9.

Sun, R. (Ed.) (2008). *The Cambridge handbook on computational psychology*. New York: Cambridge University Press.

Sun, R., & Bookman, L. (Eds.) (1994). *Computational architectures integrating neural and symbolic processes: A perspective on the state of the art*. Norwell, MA: Kluwer Academic Publishers.

Sun, R., & Ling, C. (1998). Computational cognitive modeling, the source of power and other related issues. *AI Magazine, 19*(2), 113–120.

Sun, R., Merrill, E., & Peterson, T. (2001). From implicit skills to explicit knowledge: A bottom-up model of skill learning. *Cognitive Science, 25*(2), 203–244.

Sun, R., & Naveh, I. (2004). Simulating organizational decision-making using a cognitively realistic agent model. *Journal of Artificial Societies and Social Simulation, 7*(3). http://jasss.soc.surrey.ac.uk/7/3/5.html.

Sun, R., Slusarz, P., & Terry, C. (2005a). The interaction of the explicit and the implicit in skill learning: A dual-process approach. *Psychological Review, 112*(1), 159–192.

Sun, R., Coward, L. A., & Zenzen, M. J. (2005b). On levels of cognitive modeling. *Philosophical Psychology, 18*(5), 613–637.

Sun, R., & Zhang, X. (2004). Top-down versus bottom-up learning in cognitive skill acquisition. *Cognitive Systems Research, 5*(1), 63–89, March 2004.

Sun, R., Zhang, X., Slusarz, P., & Mathews, R. (2007). The interaction of implicit learning, explicit hypothesis testing learning, and implicit-to-explicit knowledge extraction. *Neural Networks, 20*(1), 34–47, 2007.

Suppe, F. (Ed.) (1977). *The structure of scientific theories*. Urbana, IL: University of Illinois Press.

Taber, C. S., & Timpone, R. J. (1996). *Computational modeling*. Thousand Oaks, CA: Sage.

Thagard, P. (1986). Computational models in the philosophy of science. *Proceedings of the Biennial Meeting of the Philosophy of Science Association, 2*, 329–335.

Thagard, P. (2005). How to be a successful scientist. In M. Gorman, R. Tweney, D. Gooding, & P. Kincannon (Eds.), *Scientific and technological thinking* (pp. 159–171). Mahwah, NJ: Erlbaum.

Toulmin, S. (1960). *The philosophy of science*. New York: Harper Publishing.

Turing, A. (1950). Computing machinery and intelligence. *Mind, 49*, 433–460.

van Fraassen, B. (1980). *The scientific image*. Oxford, UK: Oxford University Press.

Wermter, S., & Sun, R. (Eds.) (2000). *Hybrid neural systems*. Heidelberg: Springer-Verlag.

Please cite this article in press as: Sun, R., Theoretical status of computational cognitive modeling, *Cognitive Systems Research* (2008), doi:10.1016/j.cogsys.2008.07.002