Philosophical Psychology
Vol. 18, No. 5, October 2005, pp. 613–637

Routledge
Taylor & Francis Group

On Levels of Cognitive Modeling

Ron Sun, L. Andrew Coward, and Michael J. Zenzen

The article first addresses the importance of cognitive modeling, in terms of its value to cognitive science (as well as other social and behavioral sciences). In particular, it emphasizes the use of cognitive architectures in this undertaking. Based on this approach, the article addresses, in detail, the idea of a multi-level approach that ranges from social to neural levels. In physical sciences, a rigorous set of theories is a hierarchy of descriptions/explanations, in which causal relationships among entities at a high level can be reduced to causal relationships among simpler entities at a more detailed level. We argue that a similar hierarchy makes possible an equally productive approach toward cognitive modeling. The levels of models that we conceive in relation to cognition include, at the highest level, sociological/anthropological models of collective human behavior, behavioral models of individual performance, cognitive models involving detailed mechanisms, representations, and processes, as well as biological/physiological models of neural circuits, brain regions, and other detailed biological processes.

Keywords: Cognitive Modeling; Cognitive Architecture; Level; Causality

1. Introduction

In this article we will argue for the importance of cognitive modeling, in terms of its value to cognitive science and other social and behavioral sciences, and, in turn, propose a leveled, or hierarchical, framework for cognitive modeling.

Models in cognitive and social sciences may be roughly categorized into computational, mathematical, or verbal models. Each model may be viewed as a theory of whatever phenomena it purports to capture. Although each of these types of models has its role to play, we will be mainly concerned with computational modeling, and computational cognitive architectures in particular. The reason for this emphasis is that, at least at present, computational modeling appears to be the most promising in many ways, and offers the flexibility and expressive power that no

Correspondence to: Ron Sun, Cognitive Science Department, Rensselaer Polytechnic Institute, Troy, NY 12180, USA. Email: rsun@rpi.edu

ISSN 0951-5089 (print)/ISSN 1465-394X (online)/05/050613-25 © 2005 Taylor & Francis
DOI: 10.1080/09515080500264248

other approaches can match, in terms of providing a variety of modeling techniques and methodologies, or in terms of supporting practical applications (e.g., in military simulation; see Pew & Mavor, 1998; Ritter, Shadbolt, Elliman, Young, Gobet, & Baxter, 2003).

The main point of this article is, however, the idea of a multi-level (hierarchical) approach toward cognitive modeling in cognitive science that ranges from social to neural levels. The levels of models that we conceive include: at the highest level, sociological/anthropological models of collective human behavior; behavioral models of individual performance; cognitive models involving detailed (hypothesized) mechanisms, representations, and processes; as well as biological/physiological models of neural circuits, brain regions, and other detailed biological processes.

In physical sciences, a rigorous set of theories is a hierarchy of descriptions/explanations in which causal relationships among entities at a high level can be mapped to causal relationships among simpler entities at a more detailed level. We argue that a similar hierarchy will make it possible to establish a rigorous scientific approach to cognition.

In the remainder of this paper, first, the value of cognitive modeling in studying and understanding cognition is addressed. Modeling with cognitive architectures in particular is examined. With cognitive modeling in mind, and in particular cognitive architectures as the main means for cognitive modeling, a multi-level framework is developed, which involves correspondence of causal explanations across levels, and is argued on the basis of analogy with the physical sciences. Some possible criticisms of this approach are addressed. A brief evaluation of cognitive architectures in the light of this multi-level framework follows. Finally, some concluding remarks complete this paper.

2. The Value of Cognitive Modeling

Let us discuss issues related to the importance of cognitive modeling in studying and understanding cognition. This serves as a justification for undertaking the multi-level work advocated and described in this article, in terms of the potential contributions that cognitive modeling can bring about.

There are reasons to believe that the goal of understanding the human mind/brain strictly from observations of human behavior is ultimately untenable, except for small and limited task domains. The rise and fall of behaviorism is a case in point. This case may also be argued on the basis of analogy with physical sciences, as we will do in section 5. The point is that the mechanisms of the mind/brain 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. It would be extremely hard to understand the human mind/brain in this way, just as it would be extremely hard to understand a complex computer system purely on the basis of testing its behavior, if we do not have any *a priori* ideas about the nature, the inner working, and the theoretical underpinnings of that system.[1] Therefore, by all means, theoretical developments need to go hand-in-hand with experimentation on human behavior. Given the complexity of the human mind/brain, and its manifestation in behavioral flexibility, complex process-based theories, or computational models (in the broad sense of the term), are necessary to explicate the intricate details of the mind/brain. Without such complex process-based theories or models, experimentation can only be blind—leading to the accumulation of a vast amount of data without any apparent purposes or any apparent hope of arriving at a succinct, precise, and meaningful understanding. This is descriptive of many areas of cognitive and social science research (although clearly not all).

It is true that even pure experimentalists may often be guided by their intuitive theories in designing experiments and in generating their hypotheses. So, it is reasonable to say that they are in practice not completely blind. Without detailed computational theories, however, most of the details of an intuitive theory are left out of consideration. Nevertheless, there are many reasons to believe that the key to understanding cognitive processes is often in fine details (Anderson & Lebiere, 1998), which one may claim only computational modeling can help to bring out. 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 gets trapped by the pitfalls mentioned above (even though inevitably vague intuitive theories may exist).

To further highlight the importance of computational cognitive modeling in theory building, instead of purely mathematical equations specifying details of these processes, a combination of mathematical equations and computational procedures (i.e., algorithms) should be used for cognitive modeling, as currently practiced in computational cognitive modeling. The advantages of using such extended forms are the flexibility and the expressive power such forms affords us, which many in the cognitive modeling community argue are necessary for understanding a system as complex as the human mind/brain. Pure mathematics, developed to describe the physical universe, may not be sufficient for understanding a system as complex as the human mind/brain. Compared with scientific theories in other disciplines (e.g., physics), computational cognitive modeling may be mathematically less elegant. But the point is that the human mind/brain itself is likely to be less mathematically elegant compared with the physical universe, and 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. Computational cognitive models provide a viable way (the only viable way, as some may argue) of specifying complex and detailed theories about cognition. Consequently, they may provide detailed interpretations and insights that no other experimental or theoretical approach can provide.

One particularly important strand of cognitive modeling work is based on cognitive architectures—i.e., the essential structure and process of a (broadly scoped) domain-generic computational cognitive model. They are used for a broad, cross-domain analysis of cognition and behavior (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 between modules, and a variety of other essential aspects (Sun, 1999). They help to narrow down possibilities, provide scaffolding structures, and embody fundamental theoretical assumptions. The value of cognitive architectures has been argued many times before (see, for example, Anderson & Lebiere, 1998; Newell, 1990; Sun, 2002, 2003).

They are clearly more than simulation tools, or programming languages of some sort (Newell, 1990). They are theoretically pertinent because they embody theories in a unique and, we believe, indispensable way. In fact, cognitive architectures are broad, generic theories of cognition. A cognitive architecture carries with it a set of structural (or architectural) assumptions, which form the basis of the theory that the architecture embodies. The development of details of a theory can only be done on the basis of such basic theoretical assumptions. On that basis, more detailed mechanisms are specified for various processes, such as memory of various forms, learning of various types, perception, decision-making, and reasoning. These mechanisms are also the crucial elements of the theory. These assumptions (structural or mechanistic) represent a theoretical commitment, the implication of which can then be explored. As in other forms of computational cognitive modeling, instead of purely mathematical equations, a combination of mathematical equations and computational procedures (algorithms) is used in cognitive architectures, with the advantages of being more flexible and more expressive.

In addition, some parameters may be specified uniformly *a priori* in a cognitive architecture, which form another important layer in theorizing about the mind/brain as embodied by a cognitive architecture. That is, they constitute a set of additional assumptions about cognitive processes, on top of architectural and mechanistic assumptions. The basic structural assumptions and the specifications of essential mechanistic processes may lead to the identification of values of many parameters (if such parameters have clear theoretical interpretations). Beside that, empirical data may provide means for estimating values of parameters (of course on the basis of prior theoretical assumptions).[2]

Cognitive architectures make broad assumptions, yet leave many details open. The essential considerations are often concerned with overall structural assumptions. But for practical reasons, a great many details need to be specified. They need to be specified so that (among other things) cognitive architectures can be tested and validated. During the course of the development of cognitive architectures, these details are being filled in through theoretical and/or experimental work.

3. Levels of Analysis, Levels of Modeling

With the aforementioned computational cognitive modeling approach, in order to deal with varying scales of phenomena and complexity of tasks, we believe that a *multi-level*, or hierarchical, framework should be developed. The main task of computational cognitive modeling is to develop models of cognitive agents and their interactions. Because of the difficulty and the complexity of this task, it is important to find the best way possible to approach this task. Therefore, let us begin by looking into some foundational issues.

First, what is the nature of 'scientific understanding' and 'scientific explanation', which are what cognitive modeling purports to provide? Over the last 500 years of the growth of modern science, the conception in philosophy of science has changed, as physics in particular has become more and more abstract and mathematical. For example, 'explanation' is commonly interpreted as identifying causes. This is relatively straightforward when used in the context of an example such as the heat of the sun explaining rain and snow: heat causes evaporation of surface water, which results in precipitation when the water vapor rises to higher altitudes where the temperature is lower. However, interest in a deeper definition of 'explanation' increased as many of the scientific theories of the 20th century came to construe, construct, and depend upon constructed objects (e.g., atoms, fields of force, genes, etc.) that are not directly observed by human senses.

Consequently, one important strategic decision that one has to make with respect to cognitive science is the level(s) of analysis/abstraction at which we model cognitive agents. In this regard, we need to better understand the possibility of different levels of modeling in cognitive science. Below, we will describe in some detail an abstract framework regarding different levels of analyses involved in developing models in cognitive science, which we inevitably encounter when we go from data analysis to model specifications and then to their implementations.

We note that traditional theories of multi-level analysis hold that there are various levels each of which involves a different amount of computational detail (Marr, 1982). In Marr's theory, first there is the *computational theory* level, in which we determine the proper computation to be performed, its goals, and the logic of the strategies by which the computation is carried out. Second, there is the *representation and algorithm* level, in which we are concerned with carrying out the computational theory determined at the first level and, in particular, the representation for the input and the output and the algorithm for the transformation from the input to the output. The third level is the *hardware implementation* level, in which we physically realize the representation and algorithms determined at the second level. According to Marr, these three levels are only loosely coupled, or relatively independent; there are a wide array of choices at each level, independent of the other two. Some phenomena may be explained at only one or two levels. Marr (1982) emphasized the "critical" importance of explicit formulation at the level of computational theory—i.e., the level at which the goals and purposes of a cognitive process are specified, and internal and external constraints that make the process possible are worked out and related to each other and to the goals of computation. His explanation was that the nature of computation depends more on the computational problems to be solved than on the way the solutions are implemented. In his own words, "an algorithm is likely to be understood more readily by understanding the nature of the problem being solved than by examining the mechanism (and the hardware) in which it is embodied" (p. 24). Thus, he preferred a top-down approach, from a more abstract level to a more detailed level (see Figure 1). We believe that such theories focus too much on relatively minor differences in computational tools (e.g., algorithms, programs, and implementations).

Another variant is Newell's and Simon's (1976) three level theory. Newell and Simon proposed the following three levels.

1. The knowledge level. Why cognitive agents do certain things is explained by appealing to their goals and their knowledge, and by showing rational connections between them.
2. The symbol level. The knowledge and goals are encoded by symbolic structures, and the manipulation of these structures implements their connections.
3. The physical level. The symbol structures and their manipulations are realized in some physical form(s).

Sometimes this three-level organization was referred to as "the classical cognitive architecture" (Newell, 1990). The point being emphasized is very close to Marr's view: what is important is the analysis at the knowledge level and then at the symbol level, i.e., identifying the task and designing symbol structures and symbol manipulation procedures suitable for it. Once this analysis (at these two levels) is worked out, the analysis can be implemented in any available physical means.

In contrast, in our view, the distinction borrowed from computer programming concerning 'computation', 'algorithms', 'programs', and 'hardware realizations', and their variations—as has been accentuated in Marr's (1982) and Newell and Simon's (1976) level theories—is rather insignificant. This is because, first, the differences among them are small and often fuzzy, compared with the differences among the processes and systems to be modeled (i.e., the differences among the sociological versus the psychological versus the intra-agent, etc.). Second, these different computational constructs are in reality closely tangled: we cannot specify algorithms without at least some considerations of possible implementations, and what is to be considered 'computation' (i.e., what can be computed) relies on algorithms, especially the notion of algorithmic complexity, and so on. Therefore, in actuality, we often must somehow consider together computation, algorithms, and implementation. Third, in our view, the separation of these computational details failed to produce any major insight, but theoretical baggage. A reorientation toward a systematic examination of *phenomena* instead of the *tools* we use in modeling is, we believe, a step in the right direction.

So, instead of these existing level theories, we would like to take a different perspective on multiple levels of analysis, which leads in turn to different levels of

Marr's (1982) Levels

Level	Object of Analysis
1	computation
2	algorithms
3	implementations

A New Hierarchy of Four Levels

Level	Object of Analysis	Type of Analysis	Computational Model
1	inter-agent processes	social/ cultural	collections of agents
2	agents	psychological	individual agents
3	intra-agent processes	componential	modular construction of agents
4	substrates	physiological	biological realization of modules

Figure 1 Marr's (1982) Hierarchy of Levels and a New Hierarchy of Four Levels.

modeling in exact correspondence with levels of analysis. We will sketch out a different set of levels that we have reason to believe are more relevant to cognitive science (as broadly defined). Instead of focusing exclusively, as in the traditional theory of multiple levels of analysis, on what we consider to be minor differences in the computational tools that we use (e.g., algorithms, programs, and implementations), we want to focus on the very phenomena that we are supposed to study, on their scopes, scales, degrees of abstractness, etc. Thus, in fact, the levels that we propose can be easily cast as differences among disciplines, from the most macroscopic to the most microscopic. These different levels include: the sociological level, the psychological level, the componential level, and the physiological level (see Figure 1).

First of all, there is the sociological level, which includes: collective behavior of agents (Durkheim, 1895/1962), inter-agent processes (Vygotsky, 1986), and socio-cultural processes, as well as interactions between agents and their (physical and sociocultural) environments (including, for example, cultural artifacts). This is the level that has been very much ignored traditionally in cognitive science. Only recently has the field of cognitive science acknowledged the fact that cognition is, at least in part, a social/cultural process (Lave, 1988; Vygotsky, 1986). To ignore the socio-cultural process is to ignore a major underlying determinant of individual cognition. A lack of understanding of sociological processes may result in a lack of understanding of some major structures of and constraints on individual cognition. Thus, any understanding of individual cognition can only be partial and incomplete when sociocultural processes are ignored or downplayed.[3]

The next level is the psychological level, which covers individual behaviors, beliefs, concepts, and skills. In relation to the sociological level, not only we can investigate the relations of individual beliefs, concepts, and embodied skills with those of the society and the culture, but also their changes independent of, or in relation to, those of the society and the culture (Castelfranchi, 2001). At this level, we can examine human behavioral data, compared with models and with insights from the sociological level and further details from the lower levels. Psychology (as well as folk psychology) and artificial intelligence (AI) have traditionally dealt with this level (but without a sufficient emphasis on relating it to higher and lower levels). Note that some may believe that this level is *sui generis*, and cannot be mapped to a lower level (see, for example, McDowell, 1994). We do not subscribe to such a view, on the basis of examining successful empirical work that succeeded in mapping this level to lower levels in various ways (Anderson & Lebiere, 1998; Sun, 2002).

The third level is the componential level. It is important to note that the computation of an agent is specified in terms of *components* of the agent, i.e., in terms of intra-agent processes. That is why the componential level may roughly be equated with the computational level. The concept of 'computation' can be broadly defined here, denoting anything mechanistic that can be specified as a sequence of steps and implemented on a computer (Sun, 1999). The concept involves various details, including the specification of 'computation' (in the sense of Marr, 1982, i.e., the goal of computation), algorithms, and programs.

At this level, we need to specify a computational cognitive architecture and components therein (i.e., modular structures and modules therein). In the process of analysis, we specify essential computational processes of each component as well as essential connections among various components. Thus, analysis of capacity (functional analysis) and analysis of components (structural analysis) are unified at this level. However, unlike the psychological level, work here is more along the line of structural analysis than functional analysis (while the psychological level is mostly concerned with functional analysis). At this level, we model cognitive agents in terms of components using the theoretical language of a particular paradigm (e.g., symbolic computation or connectionist networks) or their combinations. Ideas and data from the psychological level—the psychological constraints from above, which bear on the division of components and possible implementations of components—are among the most important considerations (Sun et al., 2001). This level may also incorporate biological/physiological facts regarding plausible divisions and their implementations—i.e., it can incorporate ideas from the next level down, the physiological level, which offers the biological constraints. This level results in *mechanisms*, though they are computational and thus abstract compared with physiological-level specifications of details.

Although this level is essentially in terms of intra-agent processes, computational models developed therein may also be used to model processes at higher levels, including sociological-level interactions where multiple individuals are involved. This can be accomplished, for example, by examining interactions of multiple copies of individual agents. We use computation as a means for constructing agent models from a sub-agent level (a componential level), but we may also go all the way from there to the psychological level and to the sociological level (see more discussions of mixing levels later on).

The lowest level of analysis is the physiological level (i.e., the biological substrate or implementation of computation). This level is the focus of a range of disciplines, including biology, physiology, computational neuroscience, cognitive neuroscience, etc. Although biological substrates are not among our major concerns here, they may nevertheless provide valuable input as to what kind of computation is likely employed and what a plausible architecture (at a higher level) should be like. For us, the main utility of this level is to facilitate analysis at higher levels, using low level information to narrow down choices in selecting computational architectures and choices in implementing componential computation.

Work at this level is basically the reverse engineering of biological systems. In such circumstances, what we need to do is to pinpoint the most basic primitives that are of relevance to the higher-level functioning we are interested in; while many low-level details are highly significant, clearly not all low-level details are significant or even relevant here. After identifying proper primitives, we study processes that involve those primitives, in mechanistic and computational terms.

Note that these levels as proposed above interact with each other (e.g., constraining each other) and may not be easily isolated and tackled alone. In addition, their respective territories are often intermingled, without clear-cut boundaries.

In particular, in studying cognition, there is no fixed path from either the highest level to the lowest level, or vice versa. Instead, multiple levels can, and should, be pursued simultaneously and be used to constrain and guide each other to narrow down the search for plausible cognitive architectures or interesting cognitive phenomena. Cognitive processes are too complex to be tackled in isolation; an all-out attack from multiple, converging fronts is necessary. This observation applies to levels as well as to domains.

We noticed recently that this view of levels is somewhat related to Allen Newell's ''bands'': the biological, cognitive, rational, and social bands (Newell, 1990). Newell was mainly concerned with different temporal durations of these different 'bands', from tens of milliseconds of the biological band to many hours of the social band. However, our view of levels has little to do with temporal durations of these processes, and more with different scales of phenomena and, consequently, different intellectual approaches towards studying them (as well as different causal explanations and their correspondence across levels, as will be discussed next).

In sum, we advocate that a new hierarchy of levels be adopted that focuses attention on phenomena to be studied instead of variations in the tools that we can use. Each of these newly specified levels provides a unique contribution in an integrative framework for the study of cognition. The distinction of these multiple levels is viable, and may even be necessary.

4. The Dictum of Causal Correspondence

Given this multi-level approach, we need to address details of the relationship among levels in order to fully exploit the advantage of the cascading levels. Let us review some of the characteristics of theories in physical sciences as a model for what is needed for cognitive science (and computational cognitive modeling in particular). Such theories establish hierarchies of description on many levels of detail in which causal relationships on one level can be mapped into causal relationships on other levels.

4.1. An Idealized View of Physical Sciences

Physical sciences are often taken as the paradigm for rigorous scientific theory. In these domains, theories have achieved a high degree of mathematical sophistication over the last 400 years. To explore the nature of an effective theory of cognition, it is useful to analyze the nature of theories in the physical sciences. The aim of our discussion below is to give a relatively simple (and maybe a little simplistic) view of what science aims to achieve (keeping in mind our goal of prescribing a perspective for cognitive science). More extensive discussions may be found in, for example, Lakatos (1970) and Salmon (1984).

Everyday domains of human experience are the roots from which the physical sciences have developed. Experience of movement in everyday life (e.g., hunting, warfare) has led to theories of motion from Newtonian mechanics through

Hamiltonian dynamics to special relativity and quantum electrodynamics. The ability to derive dates for planting crops from an understanding of the behavior of stars and planets led ultimately to theories of gravity, from Newtonian gravity to general relativity. The experiences of materials processing such as metalworking led to successively deeper sciences of matter, from chemistry through atomic theory to quantum mechanics and on to string theory. Deeper (more detailed) theories provide accounts that unify multiple apparently independent domains at higher levels. This unification is achieved by using small sets of relatively simple concepts that can be combined in different ways to generate the concepts and the causal relationships of the higher-level theories. In the practice of science, however, deeper (more detailed) theories do not in general replace higher-level theories (although they may result in significant revisions sometimes). For example, the deeper-level theory of quantum mechanics has not replaced the theories at the level of chemistry, and even research at the frontiers of quantum mechanics makes extensive use of classical physics (Greene, 1999).

A theory on any level creates entities at that level of descriptive detail, as well as causal relationships between those entities that correspond with a range of data. Entities at a higher level often package sets of lower-level entities in such a way that the higher-level causal relationships can be specified without reference to the internal structure of higher-level entities. However, causal relationships between detailed (deeper-level) entities must explain (or "generate") the causal relationships that exist between higher-level entities. Higher-level relationships may be generated deterministically or probabilistically from lower-level (more detailed) relationships (e.g., the behavior of gases from the statistics of large numbers of gas molecules). Whether the entities have any "absolute" reality or are simply defined by humans as an aid to understanding is immaterial to our discussion here.

Our idealized conception of a successful scientific theory has some similarities with the approach proposed by Machamer, Darden, and Craver (2000), but with some important differences. In their view, the world should be conceived as being composed of entities and activities, and mechanisms are composed of these entities and activities. Our definition of a causal relationship is essentially equivalent to their mechanism, and their definition of an activity is, in our view, equivalent to a way in which entities can interact.

An important difference however, is that in our view, a driving force for the creation of multiple levels of description is that physical scientists, like all human beings, can only handle a limited amount of information at a time. They must therefore use a high-level theory for thinking about broader phenomena, and then focus through successive levels of detail on more detailed theories. To keep the information content of description on every level within human capabilities, entities need to be arranged in hierarchies and defined in such a way that the interaction between two entities at a high level is a small proportion of the interactions within the population of more detailed entities that make up the two higher-level entities.

For example, pre-atomic or macro chemistry established entities like sulfuric acid and caustic soda, and on a somewhat more detailed level, entities like acids, alkalis, and salts. A consistent causal relationship (or law) on that level is acid plus alkali generates salt plus water. The number of different possible chemicals in macro chemistry is extremely large. On the other hand, atomic theory uses less than 100 different types of atoms and, by means of intermediate entities (like hydroxyl and hydrogen ions), can generate the acid + alkali law. However, at the atomic level, a full end-to-end description of a macro chemistry process would be long and convoluted, because it would include a description of the behaviors of all the approximately 10^{24} atoms making up a typical chemical reaction. The information density of such a description would make it too complex to be grasped within human cognitive capacities. In practice, the physical sciences establish multiple levels of description that bridge the gap between macro phenomena and fundamental theories like quantum mechanics.

Likewise, computer scientists established multiple levels of description of computation. At a higher level, a high-level programming language (e.g., an 'imperative language') is used, which has many different programming constructs for the convenient description of computation. At a lower level, a machine language (or an 'assembly language') is used, which normally has far fewer different constructs. At an even lower level, transistors are used for all operations, which have only two states (on/off). All high-level language constructs can be mapped to machine language constructs in an automated way (through a 'compiler'). All high-level language constructs, as well as all machine language constructs, can also be straightforwardly mapped to the operations of transistors. However, a description of a reasonably complex computational task based on transistors would be extremely difficult to understand. Even a description based on a machine language would be long and hard to read, due to information density. Hence, higher-level descriptions are necessary.

In the physical sciences, bridging of levels is often achieved by creating descriptions at a detailed (deep) level of small but typical processes within a high-level phenomenon, and achieving consistency with the high-level phenomenon as much as possible. Consistency between levels implies that the causal relationships between high-level entities in a process are generated by the causal relationships between groups of detailed entities making up the high-level entities. A significant inconsistency between levels would invalidate the theory. If a large and representative sample of such consistency tests all give positive results, the detailed (deep) theory is considered valid. In this situation, a correspondence exists between high and low level descriptions. In a successful theory, causal relationships at the deeper level may predict unexpected but experimentally confirmable causal relationships at higher levels.

Note that the high-level processes that actually occur are only a tiny subset of all the high-level processes that could conceivably occur given the detailed-level entities. Although any conceivable high-level process could be described in terms of the detailed entities, the subset that actually occurs is determined by the actual configuration of entities that happens to exist (in physical science terms, the boundary conditions). The identity of the subset may or may not be determined *a priori* from the properties of the individual entities alone. In this sense, some high-level processes may be "emergent".

4.2. An Equally Idealized View of Cognitive Science

A scientific theory of cognition (e.g., computational cognitive modeling), by analogy with the physical sciences, should ideally create descriptions of causal relationships starting at the level of phenomenological description of sociocultural processes and moving, via a number of intermediate levels, to a physiologically detailed description. There would be fewer types of entities and causal relationships at deeper (more detailed) levels. Higher-level entities would be made up of sets of more detailed entities, and causal relationships at a higher level would be generated by the causal relationships amongst the equivalent entities at more detailed levels. However, the causal relationships between higher-level entities could often (but not necessarily always) be described without reference to their internal structures.

The capability, at least in principle, to map collective phenomenological properties all the way down to neural properties (or other detailed level descriptions) is an essential aspect of an effective theory of cognition in a sociocultural context. In contrast, the ability of a high-level theory to accurately model high-level phenomena is a necessary but not sufficient condition for effectiveness. For example, in astronomy, the Ptolemaic theory of planetary motion based on epicycles around a hierarchy of mathematical points could account for observations at least as accurately as the Copernican model when that model was proposed. Addition of more epicycles could enable it to match observations to yet higher degrees of accuracy. However, a model based on orbits around mathematical points had much less capability for deeper-level modeling than the Copernican model, in which an orbit was generally related to the presence of an identifiable astronomical object at a central point in the orbit.

We contend that an effective scientific theory of cognition would be a "mechanistic" explanation, i.e., an explanation that "treats [the cognitive] systems as producing a certain behavior in a manner analogous to that of machines developed through human technology" (Bechtel & Richardson, 1993). What can be expected of such a theory of cognition is an appreciation of how causal relationships at a deeper level give rise to the phenomena of cognition at a higher, phenomenological level (individually or collectively), as well as a better understanding of how silicon-based systems might generate similar phenomena.

A mechanistic or process-based explanation is generally taken to be 'computational' in the broad sense of that term; here it is used to denote any process that can be realized computationally, ranging from chaotic dynamics (Freeman, 1995) and Darwinian competition (Edelman, 1989), to quantum mechanics (Penrose, 1994). In terms of the sufficiency of such computational explanations, Jackendoff (1987) proposed the following hypothesis: "Every phenomenological distinction is caused by/supported by/projected from a corresponding computational distinction"

(p. 24). Earlier, we argued for the usefulness of computational models, which we will not repeat here.

A theory of cognition that is analogous with a theory in the physical sciences should ideally first establish a set of entities or conditions, $\{C_1, \ldots, C_n\}$, at a high level with consistent causal relationships, e.g., the presence of C_1 plus C_2 causes C_3. Then, the theory must also establish a (smaller) set of entities or conditions at a lower level so that different combinations of entities or conditions, $\{c_1, \ldots, c_n\}$, correspond with the high-level set, in such a way that if C_1 plus C_2 causes C_3 at the high level of description, then at the detailed level, the combination of c_1, c_2, etc. corresponding with C_1 plus C_2 at the high level causes the combination of c_3 corresponding with C_3 at the high level. Descriptions at the high level contain less densely packed information than descriptions at the detailed level. This often means that the number of entities that are needed at the high level in general would be larger than the number of entities at the detailed level, and the number of different types of causal relationships at the high level would generally also be larger.

The lower informational density of description at a higher level means it is possible that many states at a more detailed level could correspond with the same state at the higher level. Instances of a high-level state are instances of detailed-level states (i.e., correspondence of categories across levels). However, normally, no two different states at a high level can correspond with the same state at a detailed level (Braddon-Mitchell & Jackson, 1996; Kim, 1993). Furthermore, the objects at a higher level are defined in such a way that most of the interactions at deeper levels occur within higher-level objects, and only a small residual amount of interaction occurs between objects. The higher-level model can then explicitly include only these residual interactions and the volume of information is kept within human capabilities.

In addition, note that the appropriate choice of objects and causal relationships, at any one level, is determined by the phenomena to be studied. For example, within small domains on the surface of the Earth, atoms and chemicals are useful objects. In larger domains, mineral deposits are useful, and for yet larger domains, continental plates. Similarly, in cognitive science, useful objects of analysis are identified through the confluence of a variety of considerations. These considerations may include levels of analysis, objectives of analysis, data/observations available, accuracy of data/observations available, information from adjacent levels, etc. Specifically, at a high level (the psychological level), we may identify objects of analysis such as accuracy (of a particular type of action), response time, recall rate, etc. At a lower level (the componential level), we may identify objects such as rules for action decision-making, associative memory, inferences, etc. At a slightly lower level, we may instead identify details of mechanisms such as rule encoding (for representing action rules), associative strengths (between two memory items), backward versus forward chaining (in making inferences), etc. At a very low level (the physiological level), we may identify neural correlates of action rules, associative memory, etc., as well as their details (e.g., their encodings and parameters).

Correspondences across levels are often worked out with a great deal of effort, through trial-and-error empirical work. For example, mapping neural circuits to cognitive entities (or vice versa) is a tremendously difficult and tedious process. Consistency between levels implies that the causal relationships between high-level entities in a process are generated by the causal relationships between groups of detailed entities making up the high-level entities. If a representative sample of consistency tests across levels leads to positive results, the deep theory may be considered valid.

In relation to the notion of 'hierarchies of models', we also need to examine the notion of 'modules'. In cognitive science, many different types of module have already been proposed: e.g., peripheral modules, domain-specific modules, and anatomical modules. Proposed peripheral modules include early vision, face recognition, and language production and comprehension. Such modules take information from a particular modality and only perform a specific range of functions. Proposed domain-specific modules include driving a car or flying an airplane (Hirschfeld & Gelman, 1994; Karmiloff-Smith, 1992). In such modules, highly specific knowledge and skill are well developed for a particular domain, but do not translate easily into other domains. Anatomical modules are isolatable anatomical regions of the brain that work in relative independence (Damasio, 1994). In an early discussion of cognitive modules, Fodor (1983) proposed that modules have proprietary input domains and dedicated neural hardware, and generate information in accordance with algorithms that are not shared with or accessible to other modules.

All these module definitions can be interpreted as different instantiations of the requirement for less external interactions and more internal interactions. Different types of modules are different ways to achieve minimized information exchange among different parts of a system. Minimization of information exchange is equivalent to requiring that the difference between internal and external interaction be as large as possible across all modules in a system. In particular, the concept of information hiding introduced by Parnas (1972) is that a module processes information hidden from other modules.[4]

Notice one particularly significant difference between our view of levels and that of Marr (1982). Marr argued that complete understanding of information processing in cognition required understanding on three levels: the levels of computational theory, representation and algorithm, and hardware implementation. However, Marr's position was that the three levels are only loosely coupled, since "the choice of algorithm is influenced, for example, by what it has to do and by the hardware in which it must run. But there is a wide choice available at each level, and the explication of each level involves issues that are rather independent of the other two" (p. 24). Our notion of hierarchies is rather different: The functionality described at different levels is consistent with each other. The difference is mostly in the information density (level of details) and thus the length and the complexity of description, and in the (likely) smaller number of entities needed by description at a more detailed level. However, causal relationships at one level will have to correspond (in some way) with causal relationships at other levels.

5. How Do We Proceed from Here?

In order to establish a description hierarchy based theory of cognition, the first requirement is a phenomenological description of social, cultural, and cognitive processes. Then, more detailed and more precise data are collected through a variety of means (e.g., through psychological experiments). On that basis, essential causal relationships between constructed entities, at one particular level (or across several levels), need to be identified. Based on a set of causal relationships, one may then proceed to construct a process-based description, including some key details, at a particular level (or across several levels), Then, a fully fleshed out computational model (including the specification of all needed algorithms and data structures) is developed. Finally, the computational model is implemented (in some way, on some medium), and a running program of the model is obtained. Detailed simulation data are thereby generated.

When cognitive architectures are used, however, a *two-stage* process is involved. First, essential structures and processes are constructed ahead of detailed simulation of any specific domain. The process of identifying and implementing these essential structures and processes also goes from phenomenological descriptions all the way to an implemented program. During the second stage, on the basis of an implemented cognitive architecture, the process going from phenomenological analysis to computational implementation is repeated, concerning specific details of a particular task domain. This analysis is constrained by the earlier identification and implementation of essential structures and processes in the cognitive architecture used for implementing the model of the particular task domain.

Analysis and implementation are often limited to within a particular level at a time (inter-agent, agent, intra-agent, substrate, etc.). However, this is not necessarily the case: *cross-level* analysis and modeling could be very intellectually enlightening, and might even be essential to the progress of computational cognitive modeling. For example, the cross-level link between the psychological and neurophysiological levels has been very beneficially explored in recent years (in the form computational/cognitive neuroscience; see, for example, Damasio, 1994; LeDoux, 1992; Milner & Goodale, 1995). For another example, the psychological and social levels should be crossed in many ways (and may even be integrated), in order to generate new insight into social phenomena on the basis of cognitive processes and, conversely, to generate insight into cognitive phenomena on the basis of sociocultural processes (Sun, 2001). The ability to shift freely between levels, or to understand the mapping between levels, is a critical part of scientific work.

More specifically, models for simulating (large-scale) social phenomena may be made to correspond to models of individual agents. In other words, the social level may be mapped to the psychological level in constructing models for social simulation. Such models are known as agent-based social simulation (Axelrod, 1984). Although the view of agents in such simulations tends to be extremely simple, they nevertheless illustrate a certain degree of (rough) correspondence across levels. Such cross-level analysis is useful, and these simulation models are making a significant impact on many fields of social sciences research (Axtell, Axelrod, & Cohen, 1996). (More rigorous correspondence, i.e., integration, is of course also possible, as will be discussed below.)

We note that much of the work of science takes place at higher levels. Scientists, like all human beings, can only handle a limited volume of information at one time, and must use a higher-level theory for thinking about a broader phenomenon. If necessary, they can then focus in on a small aspect of that phenomenon to develop or apply more detailed theories. For example, an engineer attempting to determine the reasons for a problem in a chemical manufacturing plant will mainly think at the macroscopic chemical level, but with occasional shifts to the atomic (or even quantum mechanical) level. Likewise, a social scientist may mainly think at the macroscopic social level, but occasionally more detailed probes into individual minds at the psychological or the componential level may be necessary in order to understand details of the cognitive basis of some social phenomena. Conversely, a cognitive scientist may mainly think at the psychological or componential level, but sometimes a broader perspective at a more macroscopic level may be necessary to understand the sociocultural determinants of individual cognition.

Beyond such cross-level analysis, there may even be *mixed-level* analysis. This idea may be illustrated by the research at the boundaries of quantum mechanics. In deriving theories, physicists often start working in a purely classical language that ignores quantum probabilities, wave functions, etc., and subsequently overlay quantum concepts upon a classical framework (Greene, 1999). This approach is not particularly surprising, since it directly mirrors our experience. At first blush, the universe appears to be governed by laws rooted in classical concepts such as a particle having a definite position and velocity at any given moment in time. It is only after detailed microscopic scrutiny that we realize that we must modify such familiar classical ideas. Thus, we view the differences and the separations amongst levels as rather fluid, and, more importantly, our idealized view does not prevent us from seeing alternative possibilities.

Another case of mixing levels is as follows. The objects and causal relationships at higher levels may be defined as combinations of more detailed objects and more detailed causal relationships. In the ideal case, the causal relationships between objects at higher levels can be defined in simple terms with 100% accuracy without reference to the internal structure of those objects as defined at more detailed levels (e.g., the internal structure of atoms at the quantum mechanical level can largely be ignored in macroscopic chemistry). In practice, however, this ideal is often not fully achieved, and the simpler causal relationships at a higher level sometimes generate predictions that are less consistent with observations than those generated at a more detailed level. A model must therefore have rules that indicate the conditions under which a more detailed model must supersede the higher-level model, or, in other words, when the generally negligible effects of the internal structure of higher-level objects must be taken into account. Therefore, again, it must be possible to integrate models on adjacent levels. We would expect similar issues to arise in cognitive modeling.

It is worth noting that Anderson (2002) provided a preliminary analysis of some possibilities in terms of mixing different levels in cognitive modeling, in relation to Newell's 'bands' (i.e., temporal scales). He showed that events of tens of milliseconds or hundreds of milliseconds could have a significant impact on learning occurring over a period of 100 hours or more. This mixing of vastly different timescales is an instance of what we have been referring to as mixing of levels.

As another example relevant to cognitive modeling, social simulation (i.e., modeling of large-scale social phenomena), despite its abstractness, may sometimes require the use of cognitive architectures (detailed models of individual cognition) in order to generate accurate predictions, and, more importantly, deep understanding of the root causes of some social phenomena. This involves mixing the social level, the psychological level, and the componential level. Sun and Naveh (2004) presented just such a case. They showed how very detailed cognitive models might be used in a high-level social simulation, namely, the simulation of organizational decision-making. Many cognitive parameters were shown to have significant effects on collective outcomes; therefore, they cannot be abstracted away in this and other similar social simulations. Mixing levels, or integrating models at different levels, is often necessary in order to improve the accuracy and predictive power of social simulations.

Finally, regardless of whether within or across (including mixing) levels, more needs to be said regarding causal explanations in science in general, and in computational cognitive modeling in particular. In this regard, we should note that modern science very much lacks a solid understanding of the notions of causality and causal explanations. One only needs to look at the contemporary discussions of interpreting quantum mechanics to see this point. Salmon (1998), in discussing causal explanations in science, sketched out the standard model of causality and claimed that it (and its variants) was profoundly mistaken and a radically different notion of causality needed to be developed.[5] Likewise, Smolin (2001) argued that in general relativity, the causal structure of events can itself be influenced by those events. The causal structure is thus not fixed. Causal connections may be spatio-temporally continuous entities that do not need to be constructed out of anything else. In general, a *process ontology* of causality was emphasized, as opposed to the standard substance ontology as in old Newtonian physics (Salmon, 1998).

In relation to these ideas, it is paramount that a better understanding of what constitutes adequate causal explanations in cognitive science—particularly in computational cognitive modeling—be developed. Salmon and other philosophers of science provided some theoretical ideas that may be further developed, for the specific purpose of laying a better foundation for cognitive science. In this endeavor, we would like to suggest that, as in contemporary physics, a shift to process ontology (as opposed to old substance ontology) is called for. Such a shift would be consistent with the major thrusts in contemporary theoretical physics and the philosophy of science (Salmon, 1998; Smolin, 2001). Cognitive architectures may well embody process ontology, because they may be structured in such a way that various factors, input, output, and internal representations, interact in complex ways (e.g., see Anderson & Lebiere, 1998; Sun, 2002).

A further point is that, as in contemporary physics, in cognitive science we can get a handle on essential causal structures and processes of cognition by looking for what is invariant under different mappings from one domain to another or from one level to another. These relatively invariant structures are the likely manifestations of a *causal nexus*—a region where essential causal structures and processes are centered (Salmon, 1998). This is exactly what some existing cognitive architectures are attempting to do. A few cognitive architectures (such as ACT-R and CLARION) are currently being applied to a variety of different cognitive task domains and to different levels as well (Anderson & Lebiere, 1998; Sun, 2002), and through such wide-ranging exploratory processes, it is likely that the essential causal nexus may be identified. For example, in the work involving the CLARION cognitive architecture, it was discovered that the mechanism governing the interaction of the implicit and the explicit processes helps to explain a variety of phenomena (Sun, 2002). Thus, it is likely that the mechanism governing the interaction of the implicit and the explicit processes lies at the center of an essential causal nexus.

6. Criticisms and Counter-criticisms

Computational modeling approaches, especially cognitive architectures, have been criticized on a number of grounds related to validation: models with lots of free parameters can fit many possible results, variability in the data may not be captured by model fitting, a tight fit may be meaningless when there is a lot of noise, a good fit may not be interesting if many models can fit the data well, etc. (see, for example, Roberts & Pashler, 2000).

Schunn and Wallach (2001) argued that these problems can be addressed by judicious use of statistical and other techniques. They show that some existing techniques for analyzing data and for conducting comparisons may be used to alleviate the problems identified by Roberts and Pashler (2000). Furthermore, obtaining a good match between human data and a computational model is not as trivial a task as Roberts and Pashler seemed to suggest—it is a difficult, and sometimes arduous, task to obtain a good match.

We would further suggest that the model match process is a process of providing a detailed (algorithmic) theory of cognitive processes involved in the data in question (in the context of other theories and data interpretations). Although these criticisms above are justified to some extent, they do not lend direct support to the assertion that computational cognitive modeling (hierarchical or not) is not important or useful. Each of these aforementioned problems needs to be addressed, and they can indeed be addressed. But the baby should not be thrown out with the bath water. There are sufficient reasons to conduct computational cognitive modeling; as argued before, computational models (especially cognitive architectures) may be the only way to represent a complex theory of the mind/brain and, therefore, are essential to the enterprise of cognitive science.

Next, let us take a different tack and look again back into the history of science, in order to address this issue in a broader context, and argue against the residue of positivist thinking as represented in some of these objections. Historically, there have been several different approaches to determining natural laws. One of the earliest approaches was inductionism. The first step is to collect numerous observations, without any theoretical preconceptions. The next step is to infer theories from the observations by inductively generalizing the data into hypothesized physical laws. The third step is to make more observations in order to modify or reject the hypothesis. The problem with this approach is that modern science (e.g., physics) includes concepts that could not be inductively derived solely from the data they explain (such as forces, fields, and subatomic particles in modern physics).[6]

Another approach is hypothetico-deductionism. In this approach, the starting point is a hypothesis or a provisional theory. The theory is used to deduce what will be observed in the empirical world, and its correctness demonstrated or refuted by actual observation. A problem with this approach is that more than one theory may be consistent with the empirical data. Tycho Brahe and Nicholas Copernicus had competing models for the solar system, but every observation predicted by one theory was also predicted by the other. A second problem is that it is always possible that new observations could topple the most established theory.

The second problem was reinterpreted into a principle by the 20th century logical positivist philosophy of science that a scientific theory is only regarded as valid as long as there are no observations that contradict the predictions of the theory. Karl Popper (1959) argued that the only criterion for a useful theory is that it makes predictions that can in principle be falsified by observations. Theories are accepted or rejected on the basis of the results of such observations. However, there are significant problems with this view of the scientific method also, because a prediction in general may depend not just on the theory being tested but also on other theories involved in modeling the system being observed or involved in modeling the observational instruments.

Thomas Kuhn (1970) addressed this problem, quite successfully in our view. For Kuhn, science proceeds in two different ways. "Normal science" is predicated on the assumption that the body of theories used is consistent and correct, and observations are collected and fitted within this current scientific "paradigm". However, new and unsuspected phenomena are sometimes uncovered, and given sufficient mismatch with the current paradigm, a crisis may develop, leading to a new "paradigm", or a new body of theories. The shift from the classical to a quantum mechanical basis for physics is a prime example of such a paradigm shift.

Computational cognitive modeling can be justified methodologically, on the basis of Kuhn's idea. According to Kuhn (1970), on the assumption that a current theory is consistent and correct, observations are collected and fitted within this current theory (or scientific "paradigm"). New, 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 of an architecture are made, and gradual validation, refinement, and reorganization of details of the architecture follow. Architectural assumptions and other commitments constitute an initial theory, which then undergoes testing through matching with experimental data. Revision and refinement are inevitable when inconsistencies and incorrect predictions are discovered, or when the model is incapable of predicting something. However, when given a sufficiently high degree of mismatch between the data and the current architecture, i.e., when revision and refinement are no longer able to accommodate problems that arise, a crisis may develop, which leads to a new 'paradigm', i.e., new architectures or even new approaches towards building cognitive architectures.

However, even with this issue of validation addressed (bearing in mind that our view is unlikely to be accepted by everyone), there may still be objections to our multi-level approach. Objections to our hierarchical framework may include the following: (1) strict correspondence is unlikely, (2) neurophysiology should go all the way, (3) lower levels are irrelevant.

We respond as follows. First, there are various kinds of causal correspondence (e.g., from strictly deterministic to highly probabilistic, and from complete reduction to looser mappings, etc.). We understand that the concept of 'causal correspondence' (the same as the concept of 'causal explanation' itself) has to be a flexible and evolving one. As science progresses, these concepts are bound to change, which we have seen plenty of evidence of in the history of modern science. Therefore, a dogmatic view of them is not warranted, and it is certainly not part of our point here.

Second, many people in cognitive science and in AI believe that low levels are simply implementational details, and as such, they have only minimum relevance to understanding high-level cognition (Marr, 1982). We disagree. As stated earlier, different levels are intimately tied together by correspondence of causal explanations across levels, as well as by mixed-level causal explanations. Such tying together is necessary for developing a deep theory in cognitive science, as opposed to a shallow, or even superficial, theory. Anderson (2002) outlined another case (an empirical case) against this view. Using examples from his tutoring systems, he showed that low-level considerations can benefit understanding at a much higher level.

In direct contrast to the idea that low levels are irrelevant, there is also the view that insists that the only way to understand the human mind/brain is through neurophysiology, which can lead directly to understanding high-level thinking as well as sociocultural processes (Churchland, 1989; Damasio, 1994; LeDoux, 1992). We disagree with this view as well. Consider what we said earlier regarding descriptive complexity and information density: because of a larger amount of details at the lower levels, it would be hard to describe high-level phenomena in a clear, convincing, and humanly comprehensible manner. That is where higher-level theories come into play: they help to describe higher-level (larger scale) phenomena with higher-level, more succinct concepts and higher-level, more succinct causal explanations. Higher abstractness at higher levels may help to reduce the descriptive complexity of theories and make them humanly comprehensible and practically useful. In the case of neurophysiology, as compared with psychology, the difference in terms of amount of detail is indeed huge.

We need to strive to avoid extreme positions at either end of the spectrum of possibilities. Clearly, a proper approach would be somewhere between these extreme positions, which requires a properly balanced outlook, taking into consideration all issues, all possibilities, and the ultimate objective of achieving a clear conceptual understanding of the human mind/brain. As stated earlier, cognitive processes are too complex to be tackled in one way only; an all-out attack from multiple fronts is necessary, and these fronts include work on all of these different levels.

7. A Quick Survey

Applying this multi-level approach to computational cognitive modeling, a hierarchy of models would be needed. A cognitive architecture should, ideally, likewise have a hierarchy of descriptions, from the most abstract to the most detailed, with consistent causal correspondence (Ohlsson & Jewett, 1997).

As mentioned briefly before, ACT-R has been successful in capturing a wide variety of cognitive data (Anderson & Lebiere, 1998). Beyond capturing psychological and behavioral data, work has been going on in capturing socio-cultural processes through simulating relatively detailed cognitive processes (West & Lebiere, 2001). Furthermore, attempts have been made to map model processes onto brain regions and neurophysiological processes. Overall, the model has shown significant promise in bridging several different levels (inter-agent, agent, intra-agent, and neural).

Also as mentioned earlier, CLARION has been capturing a wide range of data in various areas (Sun, 2002). Serious attempts have been made to address sociocultural processes through simulation using this architecture (Sun, 2004; Sun & Naveh, 2004). Some promising results have been obtained. Some attempts have also been made that map model structures, components, and processes onto those of the brain through utilizing work on biological processes of reinforcement learning (Houk, Adams, & Barto, 1995; Keele, Ivry, Hazeltine, Mayr, & Heuer, 1998; Posner, DiGirolamo, & Fernandez-Duque, 1997; Sun, 2002). Although much more work is needed, this cognitive architecture also shows promise in terms of leading up to a multi-level hierarchical theory with clear cross-level causal correspondence.

Some other cognitive architectures are also making some effort in terms of bridging across levels. For example, SOAR (Rosenbloom, Laird, & Newell, 1993) is extending into a higher level, the social level, with work on teams and other group processes. On the other hand, some connectionist models, e.g., those described by O'Reilly and Norman (2002), are making strong connections between neurophysiology and psychology. RA (Coward, 2001) is also making connections to the neurophysiological level.

However, some other existing cognitive architectures may be less promising in terms of linking across levels and establishing causal correspondence. For instance, COGNET (Zachary, Le Mentec, & Ryder, 1996) has been purely at a high level of behavioral description. Its mechanisms and processes do not translate into lower-level processes. In terms of the details of its mechanisms and processes, it is not grounded in either existing psychological theories or experimental results in a very convincing way.

In sum, a range of architectural models may be evaluated for their potential capability to support causal relationships at different levels. Several models are found to have such potential. The ability of these models to support causal relationships at different levels and the consistency of the models with sociology, psychology, and physiology need to be further strengthened. In our view, models of this type have the potential to become scientific theories with an explanatory capability analogous to theories in the physical sciences.

8. Some Final Thoughts

The importance of computational cognitive modeling, especially with cognitive architectures, should be fully recognized. As we argued earlier, this approach can be fully justified methodologically.

Within such an approach, a multi-level framework can be developed, out of the need to deal with varying scales of phenomena and complexity of tasks. First, a scientific theory of cognition requires the construction of a hierarchy of different levels with consistent causal descriptions from a low level through a series of intermediate levels to high-level phenomena. It is inadequate, for example, to only look for neural correlates of cognition or to only model sociological data without reference to cognitive factors. Second, although entire cognitive processes may, in principle, be described end-to-end in detail in terms of the activities of large populations of neurons (or at an even lower level), such descriptions would not be readily comprehensible or useful. Scientific understanding depends upon the selection of key elements of cognitive phenomena and the creation of models for such elements at appropriate levels.

It has been argued that a scientific theory of cognition requires a hierarchy of consistent causal descriptions of various phenomena on many levels from the level of socio-cultural processes to the psychological and physiological levels, in which deeper levels have higher information density and higher descriptive complexity but provide "deeper" connections amongst fewer types of entities. An important point is that a consistent description within one level alone is necessary but not sufficient to demonstrate theoretical viability. It has been argued that some cognitive models do not have the capability to support these causal relationships, and may therefore be inadequate as scientific theories. But several current models have been demonstrated to have some capabilities of this type.

Note that we are not claiming that this view on multiple levels is the definitive word on this issue. We are mostly concerned with bringing up the issue and accentuating a general outlook that we believe will benefit cognitive science in the long run. Further enhancements, additions, and revisions of these levels are certainly possible and probably inevitable.

In conclusion, we believe that we have developed a number of meaningful suggestions of possibilities for theories of cognition, based on simulations with computational models. However, much detail on these suggestions remains to be worked out. Nevertheless, we hope that these suggestions will provoke further thoughts and experimental work on various aspects of computational cognitive modeling.

Acknowledgment

This work was done while the first author was supported (in part) by Army Research Institute contract DASW01-00-K-0012.

Notes

[1] In any experiment involving the human mind/brain, there are a very large number of parameters that could influence results, and these parameters are either measured or left to chance. Given the large number of parameters, many have to be left to chance. The selection of which parameters to control and which to leave to chance is a decision made by the experimenter. This decision is made on the basis of which parameters the experimenter thinks are important.

[2] We shall note that exploring the match between a model and human data is an important means of understanding the mind/brain. Obtaining a good fit is not as trivial a result as one would believe if one does not have sufficient hands-on experience in this area. Finding a good fit often involves painstaking work. The result is a detailed understanding of what affects performance in what ways. Modeling has a lot to contribute to the understanding of the mind/brain through generating detailed, process-based matching with human data.

[3] See Sun (2001) for a more detailed argument of the relevance of sociocultural processes to cognition and vice versa.

[4] One of the reasons for modular architectures is that such architectures make it easier to modify functionality, diagnose and repair problems, and design modules relatively independently (Bass, Clements, & Kazman, 1998; Kamel, 1987).

[5] He listed three aspects of causality, which he believed to be fundamental: (i) causal processes, (ii) causal interactions, and (iii) conjunctive common causes. In the process, he developed the notions of various causal forks (interactive, conjunctive, and perfect causal forks).

[6] The emphasis in the neurosciences on physiological experiments and the suspicion of high-level theories indicate a residue influence of this approach.

References

Anderson, J. (2002). Spanning seven orders of magnitude. *Cognitive Science*, *26*, 85–112.

Anderson, J., & C. Lebiere, (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*, 123–141.

Bass, L., Clements, P., & Kazman, R. (1998). *Software architecture in practice*. Reading, MA: Addison-Wesley.

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

Braddon-Mitchell, D. & Jackson, F. (1996). *Philosophy of mind and cognition*. Oxford, England: Blackwell.

Castelfranchi, C. (2001). The theory of social functions: Challenges for computational social science and multi-agent learning. *Cognitive Systems Research*, *2*, 5–38.

Churchland, P. (1989). *A neurocomputational perspective*. Cambridge, MA: MIT Press.

Coward, L.A. (2001). The recommendation architecture: Lessons from the design of large scale electronic systems for cognitive science. *Journal of Cognitive Systems Research*, *2*, 111–156.

Damasio, A. (1994). *Descartes' error*. New York: Grosset Putnam.

Durkheim, W. (1962). *The rules of the sociological method*. Glencoe, IL: The Free Press. (Original work published 1895)

Edelman, G. (1989). *The remembered present: A biological theory of consciousness*. New York: Basic Books.

Fodor, J. (1983). *The modularity of mind*. Cambridge, MA: MIT Press.

Freeman, W. (1995). *Societies of brains*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Greene, B. (1999). *The elegant universe*. New York: W.W. Norton.

Hirschfeld, L. A., & Gelman, S. (1994). *Mapping the mind: Domain specificity in cognition and culture*. New York: Cambridge University Press.

Houk, J., Adams, J., & Barto, A. (1995). A model of how the basal ganglia generate and use neural signals that predict reinforcement. In J. Holik, J. Davis, & D. Beiser, (Eds.), *Models of information processing in the basal ganglia* (pp. 249–270). Cambridge, MA: MIT Press.

Jackendoff, R. (1987). *Consciousness and the computational mind*. Cambridge, MA: MIT Press.

Kamel, R. (1987). Effect of modularity on system evolution. *IEEE Software*, *4*, 48–54.

Karmiloff-Smith, A. (1992). *Beyond modularity*. Cambridge, MA: MIT Press.

Keele, S., Ivry, R., Hazeltine, E., Mayr, U., & Heuer, H. (1998). *The cognitive and neural architecture of sequence representation*. (Technical Report no. 98-03). Eugene, OR: University of Oregon Institute of Cognitive and Decision Sciences.

Kim, J. (1993). *Supervenience and mind*. New York: Cambridge University Press.

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

Lakatos, I. (1970). Falsification and methodology of research programs. In I. Lakatos & A. Musgrave (Eds.), *Criticism and the growth of knowledge* (pp. 51–58). Cambridge, England: Cambridge University Press.

Lave, J. (1988). *Cognition in practice*. Cambridge, England: Cambridge University Press.

LeDoux, J. (1992). Brain mechanisms of emotion and emotional learning. *Current Opinion in Neurobiology*, *2*, 191–197.

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

Marr, D. (1982). *Vision*. New York: W.H. Freeman.

McDowell, J. (1994). *Mind and world*. Cambridge, MA: Harvard University Press.

Milner, D., & Goodale, N. (1995). *The visual brain in action*. New York: Oxford University Press.

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. *Communications of ACM*, *19*, 113–126.

Ohlsson, S., & Jewett, J. (1997). Simulation models and the power law of learning. In *Proceedings of the 19th Annual Conference of the Cognitive Science Society*. Mahwah, NJ: Lawrence Erlbaum Associates.

O'Reilly, R. C., & Norman, K. A. (2002). Hippocampal and neocortical contributions to memory: Advances in the complementary learning systems framework. *Trends in Cognitive Sciences*, *12*, 505–510.

Parnas, D. L. (1972). On the criteria to be used in decomposing systems into modules. *Communications of the ACM*, *15*, 1053–1058.

Penrose, R. (1994). *Shadows of the mind*. Oxford, England: Oxford University Press.

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

Popper, K. (1959). *The logic of scientific discovery*. London: Hutchinson.

Posner, M., DiGirolamo, G., & Fernandez-Duque, D. (1997). Brain mechanisms of cognitive skills. *Consciousness and Cognition*, *6*, 267–290.

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

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

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

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

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

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

Smolin, L. (2001). *Three roads to quantum gravity*. New York: Basic Books.

Sun, R. (1994). *Integrating rules and connectionism for robust commonsense reasoning*. New York: John Wiley & Sons.

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

Sun, R. (2001). Cognitive science meets multi-agent systems: A prolegomenon. *Philosophical Psychology*, *14*, 5–28.

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

Sun, R. (2003). Desiderata for cognitive architectures. *Philosophical Psychology*, *17*, 341–373.

Sun, R. (Ed.) (2004). *Cognition and multi-agent interaction*. New York: Cambridge University Press.

Sun, R., Merrill, E., & Peterson, T. (2001). From implicit skills to explicit knowledge: A bottom-up model of skill learning. *Cognitive Science*, *25*, 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*. Retrieved June 30, 2005, from http://jasss.soc.surrey.ac.uk/7/3/5.html

Vygotsky, L. S. (1962). *Thought and language*. Cambridge, MA: MIT Press.

West, R., & Lebiere, C. (2001). Simple games as dynamic coupled systems: Randomness and other emergent properties. *Cognitive Systems Research*, *1*, 221–239.

Zachary, W., Le Mentec, J., & Ryder, J. (1996). Interface agents in complex systems. In C. Nituen & E. Park (Eds.), *Human interaction with complex systems: Conceptual principles and design practice* (pp. 35–52). Needham, MA: Kluwer.