Levels of Description and Explanation in Cognitive Science[1]

WILLIAM BECHTEL
Department of Philosophy, Georgia State University, U.S.A.

Abstract. The notion of levels has been widely used in discussions of cognitive science, especially in discussions of the relation of connectionism to symbolic modeling of cognition. I argue that many of the notions of levels employed are problematic for this purpose, and develop an alternative notion grounded in the framework of mechanistic explanation. By considering the source of the analogies underlying both symbolic modeling and connectionist modeling, I argue that neither is likely to provide an adequate analysis of processes at the level at which cognitive theories attempt to function: One is drawn from too low a level, the other from too high a level. If there is a distinctly cognitive level, then we still need to determine what are the basic organizational principles at that level.

Key words. Connectionism, symbol processing, levels of organization, reduction, mechanistic explanation.

The recent attention given to connectionist, parallel distributed processing, or neural network models of cognition has raised a fundamental question about how these inquiries relate to other attempts to explain cognitive phenomenon. The use of the three different names sometimes interchangeably and sometimes distinctively revels that there is a fair amount of disagreement as to how to answer this question. The name *neural networks* suggests that these are models of actual neural systems. For some theorists, this implies that these models are, at best, tangentially relevant to the business of cognitive science. They may be useful for studying how structures in the brain function, and since cognitive activities are realized in the brain, they may be relevant for studying how cognitive activities are realized in the brain, but that is all. Some neuroscientists, especially those who adopt the term *cognitive neuroscience* for their pursuit, however, would insist that neuroscience is in the business of explaining cognitive functions. Thus, there are some investigators who would employ the term *neural networks* who take themselves to be engaged in explaining cognitive functions. For many of these theorists it is important to develop models of how the *brain* performs cognitive functions, and they are therefore quite skeptical of the research programs commonly associated with the term *cognitive science* which have not made faithfulness to neural mechanism primary.

Other theorists, who are more likely to have come to such network models by way of psychology or artificial intelligence, tend to prefer the labels *connectionism* or *parallel distributed processing*[2]. This choice of names reflects a different conception of the enterprise. For these theorists, network models are attractive only in part because of their similarity to networks of neurons. The attraction is more due to the fact that such networks provide useful frameworks for modeling

Minds and Machines **4**: 1–25, 1994.

and proposing explanations for important features of cognitive systems such as soft constrain satisfaction, graceful degradation, and content addressable memory. In this context, connectionist models have been advanced as competitors to more traditional symbolic models which employ structured representations and operations upon them.

The term *levels* often figures prominently in discussions above the status of connectionist or network models and their relation to either neural or more traditional cognitive models, including symbolic models. For example, connectionist models are sometimes claimed to be situated at a lower level than traditional information processing models, and sometimes at a higher level than neural models (Smolensky, 1988, Bechtel and Abrahamsen, 1991). However, as I shall discuss below, the use of the term *levels* in this discussion has not always been clear. If we are to appraise the contributions of connectionism to the development of scientific inquiry into cognition, we need greater clarity on how the term *levels* is used and what is entailed by locating connectionism at a given level. The goal of this paper is to make progress on this question. In Section 1 I will consider in some detail how the term *levels* figures in a number of cognitive science discussions. While suggestive, I will try to show that these analyses are not adequate for characterizing the place of connectionist models. In Section 2 I will turn to more traditional philosophical accounts which have resulted from philosophical analyses of reduction. I will show why these too are inadequate to gain an understanding of where connectionism fits into the hierarchy of levels in cognitive science and then advance an alternative conception in Section 3. Finally, in Section 4 I will apply this conception of levels to the debate over the role of connectionist and symbol processing models in cognitive theorizing.

1. Analyses of Levels in Cognitive Science

The question of the level at which research and theorizing should be done has long been a central concern for psychology. For example, the controversy between behaviorism and mentalism can be viewed as a question of whether there is an appropriate level of analysis inside the head at which psychological models could be developed. Behaviorism did not deny that processes inside the head influenced behavior; it simply denied that these processes could be analyzed psychologically and insisted that adequate theories for psychology's purposes could only be developed by focusing on factors external to the organism such as stimuli and contingencies of reinforcement. Gibson (1979) raised comparable objections to early versions of information processing psychology, contending that there was not a level at which the internal processes could be described in terms of operations performed upon representations. One could speak of the internal system *picking up* information and *resonating* to particular environmental contexts, but for Gibson it was physiology, not psychology that was needed to explicate this activity. Modern information processing psychology, in contrast, insisted that there was just such a distinctly psychological level of analysis different from those pursued in neuroscience. Most frequently information processing psychologists and researchers in allied disciplines such as artificial intelligence took this level to be characterized in terms of symbolically represented information and operations performed on these representations.

One of the most sophisticated analyses addressing the question of the relation of information processing levels to neuroscience levels was advanced by David Marr (1982). Marr proposed three distinct levels of analysis which he termed the *computational*, *representational and algorithmic*, and *implementational*. As shown in Figure 1, for Marr the computational level specifies the function that the cognitive system is to perform, the representational and algorithmic level specifies the procedures by which this is to be carried out, and the implementational level specifies the physical mechanisms which carry out this process. Marr himself approached the cognitive task of interest to him, vision, from the background of work in neuroscience. He begins his book on vision with an overview of some of the early successes neuroscientists had in identifying mechanisms involved in vision including Barlow's (1953) demonstration of ganglion cells in the frog's retina which serve as "bug detectors", Hubel and Wiesel's (1962, 1968) demonstration of edge detectors in cats and monkeys, and his own work (Marr, 1969) demonstrating capacity of Purkinje cells in the cerebellar cortex to learn motor patterns. These research endeavors suggested that merely by recording cell activities in various parts of the brain we could learn how the brain performs various cognitive functions.

Marr observes that this program failed to make further progress in the 1970s. In

Computational Theory	Representation and Algorithm	Hardware Implementation
What is the goal of the computation, why it is appropriate, and what is the logic of the strategy by which it can be carried out?	How can this computational theory be implemented? In particular, what is the representation of the input and output, and what is the algorithm for the transformation?	How can the representation and algorithm be physically realized?

Fig. 1. Marr's three levels at which any machine carrying out information processing must be understood. Adapted from Marr (1982), p. 25.

part this was due to the failure to discover more centers to which cognitive functions could be localized. For Marr, though, the more significant problem was that simply knowing that particular cells in the brain are responsive to particular sensory information does not yet reveal how these neurons contribute to vision. He argued that what was required was a different *level* of understanding:

> There must exist an additional level of understanding at which the character of the information-processing tasks carried out during perception are analyzed and understood in a way that is independent of the particular mechanisms and structures that implement them in our heads. This was what was missing – the analysis of the problem as an information processing task. Such analysis does not usurp understanding at the other levels – of neurons or of computer programs – but it is a necessary complement to them, since without it there can be no real understanding of the function of all those neurons (Marr, 1982, p. 19).

As we have already noted, Marr introduced two additional levels. At the representational and algorithmic level the inputs and outputs of the system are understood as representations and an algorithm is advanced that specifies the transformations that must be performed to generate the output representation from the input. Finally, there is the level that Marr, misleadingly, calls the *computational theory*. The name is misleading since at this level the researcher is not concerned to explain the computational procedures, but rather to specify the task to be performed by the computation system, why that task is to be done, and the constraints the task itself imposes on the performance of that task. Marr's conception of levels is considerably richer than that which figures in most accounts of psychological inquiry, which generally focus on one level. In fact, it offers a rather interesting integration of very different sorts of enterprises. For example, endeavors such as those of J.J. Gibson and other ecologically oriented psychologists might be construed as providing a theory at the computational level, while information processing accounts would offer a theories of representations and algorithms, and neuroscience would provide accounts of implementation.

Marr's analysis of levels was marshalled by David Broadbent (1985) in his criticism of connectionism. Broadbent contended that connectionist analyses are only appropriate to Marr's implementational level, but that psychological analyses are situated at Marr's computational level. Thus, Broadbent claims that connectionism is irrelevant to psychology. As Rumelhart and McClelland (1985) make clear in their response, Broadbent's discussion leaves out the intermediate level from Marr's analysis, the representational and algorithmic level. For almost all psychologists who claim to be giving a cognitive analysis this is a serious omission, for this is the level at which traditional symbolic information processing accounts are framed. One of the most common critiques of connectionism is that it errs precisely in not having the resources to offer a proper analysis at this level (Fodor and Pylyshyn, 1988). Rumelhart and McClelland contend, moreover, that the algorithmic level is precisely the level at which connectionist accounts are properly located:

> We believe that our proposal is stated primarily at the algorithmic level and is primarily aimed at

> specifying the representation of information and the processes or procedures involved in storing and retrieving information (p. 193).

In addition, they claim that accounts at all three levels are pertinent to psychological theorizing and are implicit in their research program. They do not elaborate, however, on what computational theory is assumed in connectionist modeling, but presumably it is not much different from that which is assumed in more traditional information processing accounts. They do suggest that the choice of connectionist models at the representational and algorithmic level is influenced by the need to implement the representations and algorithms in a neural architecture and that connectionist models are more capable of that than traditional symbolic information processing models. This implementation, however, is an additional activity beyond constructing connectionist models.

It may initially seem surprising that Rumelhart and McClelland place their connectionist models at the same level in Marr's hierarchy as traditional information processing models even though in their two volume account of connectionism they refer to it as offering a theory of the *microstructure* of cognition (Rumelhart *et al.*, 1986, and McClelland *et al.*, 1986). But they go on to develop their analysis by claiming "there is more twixt the computational and the implementational than is dreamt of in Marr's philosophy" (p. 195). They propose that traditional information processing accounts and connectionist accounts occupy different intermediate levels between computational theory and neural implementation, with connectionism situated beneath the traditional information processing account. This proposal prompts two further questions: how are these levels distinguished and how are they related to each other? To indicate their answer to both of these questions Rumelhart and McClelland draw an analogy to the relation between Newtonian mechanics and quantum field theory. For them, the important feature conveyed by this analogy is that the higher level theory (Newtonian mechanics) provides an *approximate* account of the phenomena for which the lower-level theory provides a more accurate account:

> Through a thorough understanding of the relation between the Newtonian mechanisms and quantum field theory we can understand that the macroscopic level of description may be only an approximation to the more microscopic theory. Moreover, in physics, we understand just when the macro theory will fail and when micro theory must be invoked. We understand the macro theory as a useful tool, by virtue of its relations to the micro theory. In this sense, the objects of the macro theory can be viewed as emerging from interactions of the particles at the micro level (p. 196).

This passage raises an interesting tension which in fact runs through much connectionist discourse and to which we will return. On the one hand the passage indicates that the phenomena discussed in macrolevel theories are *emergent phenomena*. In many accounts, the term *emergent* is used to suggest that an underlying system gives rise to phenomena which then obey laws at a higher level. The higher level is not an approximate account, but captures the real entities that result from lower level interactions. But given Rumelhart and McClelland's construal of the macrolevel as only offering an approximation, a *useful formal* *tool*, the suggestion is that the phenomena of the upper level are not real. In philosophical parlance, the upper level theory does not describe real phenomena, but is only instrumental.

Paul Smolensky (1988) has offered an analysis that is similar to Rumelhart and McClelland's, but rather more elaborate. Smolensky actually advances a pair of distinctions, contrasting on the one hand the symbolic with the sub-symbolic *paradigms* and on the other the conceptual and subconceptual *levels*. The term *paradigm* is used to designate very general theories distinguished by the nature of the entities designated in those theories. Thus, the symbolic paradigm analyzes relations between symbols that have semantics and are operated on by syntactical rules. The sub-symbolic paradigm, in contrast, describes different sorts of units, nodes in connectionist networks, which are not operated on by syntactic rules but behave in a manner characterized by mathematical laws. The prefix *sub* in *subsymbolic paradigm* is used to indicate a part-whole relations between symbols and subsymbols.

> The name "subsymbolic paradigm" is intended to suggest cognitive descriptions built up of entities that correspond to *constituents* of the symbols used in the symbolic paradigm, and they are the activities of individual processing units in connectionist networks. Entities that are typically represented in the symbolic paradigm by symbols are typically represented in the subsymbolic paradigm by a large number of subsymbols (p. 3).

The distinction between symbols and subsymbols, while part of the contrast between paradigms, brings the notion of level into Smolensky's account. Since subsymbols are components of symbols, they occupy a lower level. The distinction between the conceptual and subconceptual levels must be kept separate from this distinction between paradigms for Smolensky, since he wants to allow that models from either paradigm can be analyzed at either level. Jay Rosenberg (1990) has tried to picture the dual distinctions Smolensky advances. As Figure 2 suggests, the two paradigms are both supposed to account for cognitive phenomena. The subsymbolic paradigm clearly is supposed to operate on two levels, but Rosenberg

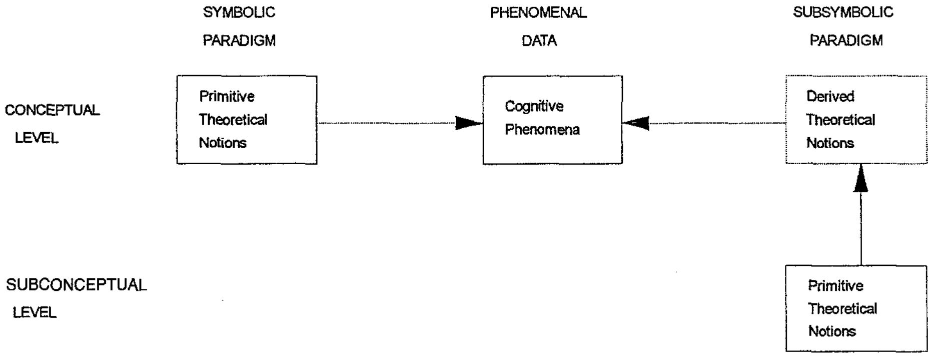

Fig. 2. Rosenberg's (1990) portrayal of Smolensky's (1988) account of the relation between the conceptual and subconceptual level on the one hand and the symbolic and subsymbolic paradigm on the other hand.

leaves a blank at the subconceptual level for the symbolic paradigm. This reflects that fact that while for Smolensky there are a variety of levels below the conceptual level in terms of which a model of the symbolic paradigm must be implemented, none is appropriately singled out as *the* subconceptual level.

It is difficult, though, to determine what exactly Smolensky means by a level. His appeal to levels is most naturally understood in terms of the part-whole relations that exist between the referents of symbols and subsymbols. Subsymbols designate features of objects, not objects themselves. Thus, in Smolensky's favorite case, which he took over from Pylyshyn, a cup of coffee might be represented in terms of a collection of features such as: upright container, hot liquid, porcelain curved surface, burnt odor, brown liquid contacting porcelain, finger sized handle, and brown liquid with curved sides and bottom. But this is trickier than might first appear. When we consider the representations themselves, it is not obvious that we should treat the representations of these features as being at a different level from the representation of the object, the cup of coffee, to which they are attributed. Notice that insofar as we have words for features, in the symbolic paradigm we might also have symbols for them. In fact, a sentence often predicates such a feature of an object: "The cup of coffee has a burnt odor." But perhaps the problems lies in the example, which designates features in terms of items we might predicate of an object. Many of the features used in the subsymbolic paradigm might better be termed *microfeatures*, for they will generally be constituents of the objects which are not identified with words in the same discourse as we refer to the object. Thus, when Rumelhart and McClelland (1986) introduce microfeatures in their model of past-tense formation they identify phonemic characteristics such as roundedness, frontalness, and backness, which designate aspects of the process by which the sound is actually produced. Even more significant is the fact that in connectionist simulations in which the representational function of a unit is a product of its learning, not the design imposed by the theorist, the units do not end up representing anything that is precisely characterizable in natural language. In Hinton's (1986) network which learns relationships in a family tree, some of the hidden units can be interpreted as representing such things as the generation from which the person comes. But while this characterizes a major portion of what the unit is responding to, the unit becomes active in other cases in which this microfeature is not present, and is off in others in which it is present. So the linguistic characterization of what the unit is representing is only approximately correct: the unit is actually functioning in a more subtle manner.

If we unpack the notion of level in terms of the semantics of the representations used in the two paradigms, however, we are not clearly differentiating levels in terms of the *mechanisms* postulated in the two theories. Some connectionist models employ units to serve the same representational functions as symbols in symbolic accounts, and there is no reason a system using a symbolic architecture could not have its representations designate the referents of subsymbols. Levels only seem to enter when we ask how representational structures are to be interpreted, not when we examine the representational structures postulated in the two theories. For Smolensky, moreover, the notion of level seems to be very much secondary to the competition between paradigms. It becomes tempting to dismiss the analysis of levels from Smolensky's discussion altogether and construe it just as an account of competition between paradigms. But if we do this there will be aspects of Smolensky's analysis we will not be able to accommodate. He phrases much of his discussion in terms of the relation between macrotheories and microtheories. He invokes as an example the relation between Newtonian mechanics and quantum mechanics, the same example cited by Rumelhart and McClelland. Moreover, Smolensky's analysis is not simply one of competing accounts that are incompatible or incommensurable with each other. He wants to allow drawing connections between the two frameworks:

> The picture that emerges is of a symbiosis between the symbolic and subsymbolic paradigms: The symbolic paradigm offers concepts for better understanding subsymbolic models, and those concepts are in turn illuminated with a fresh light by the subsymbolic paradigm (p. 19).

The notion of levels once again seems to be playing a central role: The symbolic and subsymbolic analyses are presented as describing processes at different levels in nature such that the processes identified at one level can inform the analysis at the other. Subsequently, Smolensky invokes Chomsky's competence/performance distinction to characterize the relationship: the symbolic model may provide an account of what competence in a domain would involve, while the subsymbolic connectionist account will provide a more accurate picture of actual processing that in certain conditions will exhibit that competence. The idea seems to be that in terms of a symbolic theory appropriate to the conceptual level we can characterize a competence that is realized by a mechanism operating at a different level.

Rosenberg (1990) finds there to be deep tension in Smolensky's account at just this juncture. He points out that there are two ways to analyze the relation between levels, depending upon the way one construes theories. If one takes an instrumentalistic approach to theories, theories are simply tools for predicting phenomena. The posits of the theory, however, are not given any further role. In particular, they are not taken to be real structures in the architecture of the world. From the instrumentalist's perspective, one can compare two theories with respect to how well each accounts for the phenomena and make such judgments as that one accounts for the data more precisely than the other, but that the other does give an approximate account of the data. On the realist interpretation, in contrast, there is an ontological commitment to the theoretical posits of the theory (what Rosenberg calls the *intentional content* of the theory). If these theoretical commitments turn out to be false, then there is no explanatory value to the theory for the realist. This is an absolute judgment, not a matter of degree:

> It is important in appreciating the difference between the instrumentalist and realist understanding of

> theories to recognize that, unlike descriptive fit, which is extensional and comes in degrees, *explanatory success* is intentional and not a matter of degree (p. 169).

Rosenberg argues that Smolensky is committed to a realist interpretation of theories. Otherwise the conflict between the paradigms would disappear because Smolensky acknowledges that we can always develop models in symbolic terms that perform the same as connectionist models and vice versa. Since, on purely instrumentalistic grounds, there is no basis for conflict, Smolensky can only make a case for conflict between paradigms if he adopts a realist view. But if he adopts a realist perspective, Smolensky must make a choice, according to Rosenberg. Since he takes the connectionist framework to be the correct framework for developing models of cognition, he must either construe it as a replacement for the symbolic account, and hence as a version of eliminativism, or as providing an implementation of the symbolic account, and not incompatible with it. There is no middle ground so long as we take the symbolic and subsymbolic paradigms to be explaining the same thing.[3]

There may be a way of overcoming Rosenberg's objection, but in order to explore this it is important to note that the notion of levels we are now dealing with is different from Marr's notion with which we began. Marr was concerned with *levels of analysis*; his distinction of levels does not have any ontological import. (See McClamrock, 1991, who also claims that Marr's levels are primarily concerned with different types of analyses, or what he calls *perspectives*, and that this is wrongly conflated with the ontological notion of level of *organization*.) With appropriate modification of Marr's notions of computation and of representation and algorithm, one could invoke Marr's three levels of analysis at any ontological level of organization in nature. For example, one can invoke three such levels in the analysis of an intracellular physiological process such as oxidative phosphorylation. At one level of analysis one can ask about the purpose oxidative phosphorylation serves and what constraints this purpose places upon the process itself. At another level one can seek an account of the metabolites that enter into or leave the oxidative phosphorylation process and a flow chart characterizing the transformations between input and output substances. Finally, at a third level one can seek the constitution of the actual physical components that implement that process, the enzymes that promote the chemical reactions and the membrane over which ion gradients are established to promote ATP synthesis.

Rumelhart and McClelland, Smolensky, and Rosenberg, however, are not concerned simply with levels of analysis but with *levels of structure* in nature. These structures can enter into composition relations with each other. One might build one structure out of others. This is apparent in both the use of the quantum mechanics/Newtonian mechanics analogy and the use of the terms *emergent* by Rumelhart and McClelland. In their view, Newtonian systems are composed of quantum mechanical systems, and emergent phenomena result from interactions of lower level phenomena. Thus, to clarify further the relation they envisage we need different tools than Marr offers. One place to which it might seem appropriate to turn is the philosophical literature on theory reduction, for what the philosophers who developed models of theory reduction presented themselves as doing was showing the relationships that hold between inquiries developed at different levels in the ontological hierarchy of nature (e.g., between theories of molecules and theories of atoms, between theories of living systems and theories of physical and chemical systems, and between theories of psychological phenomena and theories of biological phenomena).

2. The Philosophical Framework of Theory Reduction

The analysis of theory reduction has its origins in an account of science that begins with what are taken to be the products of scientific inquiry, namely, theories. These theories are construed as linguistic entities, and are frequently rendered in the form of axiomatic systems in which a variety of laws are construed as theorems derived from basic axioms. The function of these laws is to predict and explain phenomena in the domain of the theory. The structure of explanations and prediction is in fact taken to be identical: both predictions and explanations involve the derivations of descriptions of phenomena from one or more laws together with some pre-existing empirical conditions, known as *initial conditions*. Thus, an explanation involves having the right kind of logical relationship between sentences.

The theory reduction model simply extends this relationship to account for relations between theories. One theory is presented as explaining another theory when the statements of the second theory (e.g., laws relating temperature, pressure, and volume in a gas) can be derived from those of the first (which characterize the behavior of molecules). Two other components complete this picture. Since the vocabulary of the second theory is likely to be different from that of the first, bridge laws are required which identify terms of the reduced theory with those of the reducing theory. For example, temperature is equated with mean molecular energy. Finally, the reduced theory will only apply over a part of the domain of the first theory, so boundary conditions are required. For example, the gas laws only apply to molecular systems which constitute gases, so the boundary conditions will specify that only when we are dealing with gases can we derive the laws of temperature and volume from those of molecular motion. (See Nagel, 1960, for the classical account of theory reduction. In Bechtel, 1988, I provide a general introduction to the theory reduction model and the criticisms that have been leveled against it.)

The schema of theory reduction might seem to offer a way to account for the relation of connectionism to traditional symbolic information processing theories. This would require showing how traditional symbolic theories can be derived from and thereby reducible to those of connectionism. Those who propose that the function of connectionist theories is to show how symbolic processing might be implemented in the brain might accept such a picture.[4] But this conception cannot be accepted by connectionists who view connectionism as being in competition with symbolic theories in virtue of offering incompatible accounts of cognition. Two mutually inconsistent but internally consistent theories cannot be derived from one another.

Advocates of the theory reduction model, however, developed a modification of their view that might seem to handle this situation. Initially, the theory reduction model seemed to apply both to relations between lower level and higher level theories and to those between successor and predecessor theories. The reason for joint application is that lower level theories often were developed after higher level theories. They also seemed to correct higher level theories. Then there might be an attempt to show that the new, lower level theory could explain how the older, higher level theory could work as well as it did by showing that under special conditions, such as limits, one could derive the higher level theory from the lower level one. There is, as Nickles (1973) pointed out, a crucial difference in direction in the two cases. We speak of the higher level theory being reduced to the lower level theory, but we speak of the newer theory reducing to an older theory, especially under limit conditions. Thus, in one case we reduce a higher-level specific theory to the more general lower level theory, but in the other case we reduce the more general contemporary theory to the older theory, shown to apply only in specific circumstances.

We have already identified a more serious problem with so linking the two senses of reduction: the new theory is in conflict with the older theory. The theory reduction model used only deduction to relate two theories, but one cannot derive one set of propositions from another with which it is inconsistent. The reference to deriving the older theory under a limit does not really solve this problem since the kind of functions the theory reduction model was designed to accomplish (e.g., unifying different scientific theories and their ontologies) cannot be accomplished by having derivations only go through in limit conditions.

Fundamentally, two activities are being collapsed in the two uses of reduction. We have an *interlevel* reduction of a higher level theory to a lower level theory and what is typically an *intralevel successional* reduction of a new theory to an older one (McCauley, 1986). The latter activity does not require derivation of one theory from another, but a demonstration of some sort of similarity. This may involve similarity in the basic set of laws, but more likely simply a demonstration of why the laws of the older theory were as accurate in their predictions as they were. Generally this requires showing that some of the explanatory structure of the old theory can be recovered from the new theory under such operations as taking a limit. For example, we can derive Newtonian laws of motion from relativistic laws if we assume that velocities are low.

While recognizing that interlevel reduction and successional reduction are different, some theorists have tried to develop a common model to handle both.

Schaffner (1967), therefore, developed a more comprehensive model of the theory reduction process wherein a higher level theory is replaced by a more adequate higher level theory that shares important similarities to the old one, and this new higher level theory is then derived from the lower level theory. As Figure 3 makes clear, deduction is reserved for the interval relation, and a demonstration of similarity is employed for the successional relation.

Employing Schaffner's model, we can begin to see what was going on in Smolensky's analysis discussed in the previous section (Figure 4). Smolensky's two paradigms represent old and new theories. These are never directly derived from one another. At best they can be shown to enjoy significant similarities under a variety of conditions. The most plausible example is that both can be shown to provide similar accounts of conscious, rule-based reasoning. However, the two theories are not at the same level. So from the new, subconceptual, subsymbolic theory a new conceptual level subsymbolic theory must be derived. Under appropriate conditions, for example, when the network is engaged in reasoning from exemplars, or in solving problems by explicitly applying rules, the

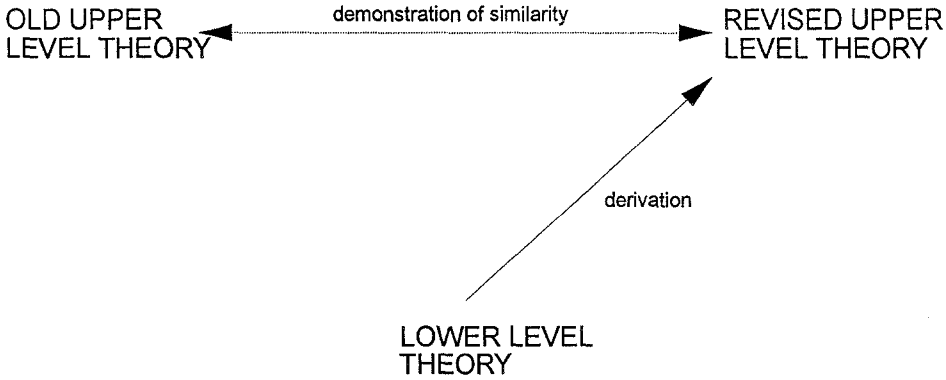

Fig. 3. Schaffner's (1967) model of theory reduction with revision of higher level theory.

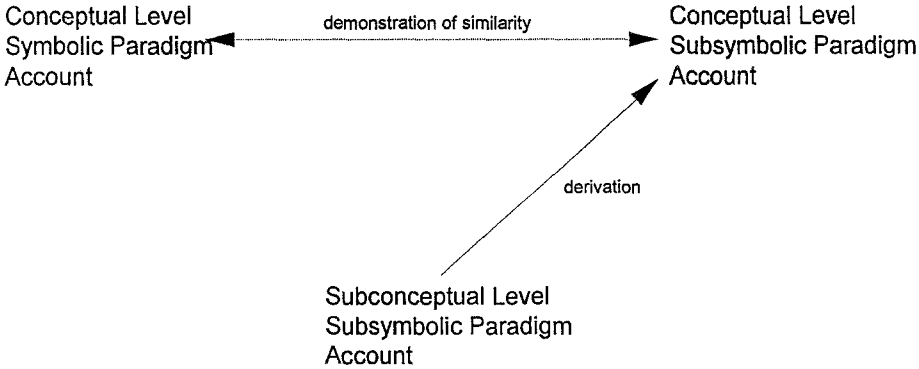

Fig. 4. Application of Schaffner's model of theory reduction and replacement to Smolensky's account of the relation of the conceptual and subconceptual levels and the symbolic and subsymbolic paradigms.

conceptual level subsymbolic theory will closely correspond to the symbolic theory. Notice that from this perspective, the symbolic theory is taken to be discredited and eliminated. In Rosenberg's terms, the symbolic theory gives no explanation at all. On the other hand, the new conceptual level theory is precisely implemented by the subconceptual level subsymbolic theory. If this account can be filled in, the tension that Rosenberg identified may be resolved. What will be accepted as real, and hence capable of offering explanations, will be the conceptual level, subsymbolic theory. The conceptual level, symbolic theory will not have any explanatory power, although insofar as we can identify similarities between it and the conceptual level, subsymbolic theory, we will be able to understand why it should have seemed like a plausible theory. What this requires, though, is a new conceptual level subsymbolic theory and as of yet we have not been offered an analysis of what this theory looks like.

In the next section I will suggest that for all of its tidiness, this framework is not sufficient to deal with the relations between levels in real science and so fails also to account for the relation between connectionist and symbolic theories. But before developing my objections, there is a feature of the theory reduction analysis worth emphasizing. The notion of level as used in the theory reduction account is often thought to correspond to ontological structures in nature (atomic interactions form one level, molecular interactions constitute another, and interactions between microscopic biological structures such as membranes and organelles yet a third). But in most accounts of the theory reduction model, levels are actually characterized only in terms of linguistic structures, the theories that provided the premises or conclusion of a deduction. Levels are propositions in a deductive hierarchy, not structures in nature. Any sense of what kind of constitutive relationship might hold between phenomena at different levels in nature is lost.

This permits a theorist great latitude in modifying theories so as to accomplish a reduction. It is conceivable that an existing lower level theory might not support the derivation of a higher level theory even though there is no inconsistency between the two. There may be laws in the higher level theory that are not deductive consequences of the laws currently stated in the lower level theory. Within the framework of the theory reduction model, it is a legitimate option to revise the lower level theory to provide it with the resources for carrying out the derivation of the lower level theory. These revisions might be motivated not by anything studied by the lower level science, but solely by the desire to be able to reduce independently confirmed higher level laws. If this is correct, it is less reasonable to see these theories of the lower level science or about lower level phenomena, but rather as theories constructed specifically to account for the higher level theories. Thus levels, construed as organizational levels in nature, once again seem to have dropped out.

(One theorist within the theory reduction framework who does seriously employ the compositional notion of level is Causey (1977). In Causey's analysis, fundamental laws of the lower-level science characterize parts of the entities described in the laws of the higher-level science in their *unbound* condition. Before relations between the theories are established, it is necessary to derive laws within the lower-level science characterizing the behavior of the compound entities which will then be equated with the entities identified in the higher-level laws. This is a demanding requirement; one that, if it can be satisfied, will make the theory reductionist claim very powerful. It restricts the lower-level theory to information about how basic lower-level entities behave outside the compound state and requires that from this and information about the structure of the compound we be able to derive information about the behavior of the compound. The problem is that this demand is so strong that there is reason to doubt if it can be met in the course of conducting actual scientific inquiry. The reason is that it is doubtful that the crucial properties about how components behave in compounds can be identified in their *unbound* condition; these properties may only be revealed when the components are joined into compounds.)

3. An Alternative Conception of Levels and Interlevel Relations

The theory reduction model was constructed largely to serve philosophical objectives, for example, showing the unity in the products – theories – generated by different sciences. Perhaps not surprisingly, therefore, it does not serve the ends of scientists, who are generally not dealing with completed theories but rather are attempting to develop adequate theories of particular domains of nature. Far more relevant for these purposes are what Darden and Maull (1977) call *interfield theories*. These are theories that attempt to draw connections between phenomena studied in different fields, often at different levels, to answer pressing questions. One context in which such questions arise is when scientists have identified a system that performs an activity of interest within the constraints of a particular environment and want to determine what processes transpire within the system that enable it to perform that activity. Answering this question requires taking the system apart to identify its components and their activities and determining the nature of the interactions between the components. (The analysis presented here is developed more completely in Bechtel and Richardson, 1993.)

In pursuing this task scientists begin to build a *model* of the system. A model is not an actually functioning system, but an account of what the components of the system are and how they are put together. The model is intended to show how the system is able to carry out whatever task it performs. As Rosenberg (1990) puts it, a model consists of elements and operations performed on or by these elements. It is worth noting that while a model may be partially described in language, most models are sufficiently incomplete and so dependent on idealizations that one cannot logically deduce from the description of the model how the model will behave. Rather, the scientist's understanding of a model tends to be more allied with perception and imagination: the scientist imagines the outcomes as different parts perform their functions and interact with each other. This is very much the way we go about understanding mechanical systems produced by human engineering. Thus, we understand how a car engine operates by imagining a flammable fluid flowing through a tube, being vaporized, compressed, and ignited to push a piston, which we envision as transferring its motion to the wheels through a set of shafts and gears. We do this without being able to complete this account so as to derive logically the characteristics of the engine's behavior. Our ability to understand this sketch depends on our previous familiarity with processes of a similar kind. If one had never witnessed an explosion, for example, it would be much more difficult to understand this account. Much the same sort of thing goes on when, for example, one explains a physiological process such as fermentation. Someone seeking an understanding of fermentation generally comes to it with a previous familiarity with certain basic chemical reactions and methods for producing and studying them. An explanation consists in showing how a sequence of these reactions can lead from sugar to alcohol, along with some account of what sorts of things (catalysts or enzymes) can initiate these reactions. Such an account is then tested by such means as producing evidence that the postulated processes are at work in the system (e.g., by showing that living cells are capable of catalyzing a reaction that figures in the model of the reaction).

Since this approach thinks of the phenomena to be explained as the product of complex systems that constitute mechanisms, it is useful to think of this as a model of *mechanistic explanation*. This view of mechanistic explanation puts a different perspective on the way we conceive of levels and relations between levels in science. Insofar as an important step in developing an explanation involves decomposing a system into its parts, we are led to focus on the compositional relations involved in nature. Models posit parts of a particular sort within a system and operations performed by these parts. The task in developing a model is to identify these parts and ascertain what they do and how they interact to produce the phenomenon of interest. Typically, a part found within a system will interact with other parts of roughly the same magnitude, and this cluster of interacting parts will constitute a level (Wimsatt, 1976).

It should be noted that finding parts and the levels at which they reside is not always an easy task. Natural systems typically do not reveal their parts when operating smoothly and so various research strategies are required to disrupt them in ways that reveal their components. Moreover, it is sometimes possible to decompose a system at high or too low a level and miss the level at which interactions transpire that are crucial to accounting for the phenomenon in question. This, for example, happened in research on oxidative phosphorylation in cells: as a result of the success in explaining fermentation in terms of reactions between enzymes and other molecules that could be extracted from cells, researchers sought to explain oxidative phosphorylation in the same manner, missing the higher-level processes involving membranes that the chemiosmotic hypothesis (Mitchell, 1966) showed to be crucial (Bechtel, 1993b).

In this account of mechanistic explanation, levels take on a far more salient status than in the traditional philosophical account of explanation. Explaining a phenomenon is in part a matter of finding the correct level for understanding particular interactions. Moreover, the resulting model is inherently interlevel. Explaining how a system works involves not only determining how it interacts with other systems, but what its components are, what they do, and how they are integrated so as to enable the system to perform the activity of interest. It should be emphasized that in developing such mechanistic explanations, scientists do not generate anything resembling a theory reduction. They do not develop two theories and connect them with deductions, but rather models that cross levels or interlevel theories (Bechtel, 1988).

4. Applying the Alternative Conception of Levels to Mental Phenomena

The account of mechanistic explanation and of the role of levels of organization in these explanations that I have been developing has been motivated in large part by attending to the biological sciences and not to psychology. Let us consider whether this approach can provide us any insight into the relation between symbolic and connectionist models of cognition. For many people, it makes no sense to attempt to explain mental phenomena by taking a system apart and finding out how it works. That is, mental phenomena do not lend themselves to mechanistic explanation. This reaction is probably largely a vestige of Cartesian dualism, which has led us to think of mental phenomena in isolation from any underlying mechanisms. Even as information processing psychology developed analyses of human cognitive performance that involved apparently mechanical procedures for manipulating information, both information and procedures for manipulating it seemed to be rather disembodied. But a naturalist would be inclined to view cognition as the product of an embodied system in the physical world and to show how cognitive activities were generated by that system. So let us see if we can employ the model of mechanistic explanation developed in the previous section to mental phenomena.

It may seem inevitable that if we are to pursue the sort of research enterprise I have been outlining with respect to mental phenomenon that we must turn to the brain and the neurosciences which investigate processes in the brain. Only in the brain, it might be thought, can we hope to discover parts in terms of which we might explain cognitive performance. While I certainly would not want to eschew information from neuroscience whenever it might be useful in developing such mechanistic accounts, neuroscience is not the only place to begin to decompose the process of cognition. Biochemistry made great headway in developing mechanical accounts of various metabolic processes when it was still impossible to determine the physical character of many of the enzymes proposed in such accounts. These enzymes were identified functionally in terms of what reactions they catalyzed and the ways in which they could be inhibited or destroyed. Thus, one can identify components functionally in terms of the processes they foster or carry out without yet being able to identify them structurally. A similar program is already well advanced within psychology as well. By defining very specific tasks that are likely to invoke what are taken to be a specific set of cognitive capacities (e.g., tasks involving memory recall, or visual analysis, or word identification) and then using behavioral measures (e.g., priming, reaction time, error patterns), psychologists are seeking to identify different cognitive processing mechanisms and the procedures each uses.

As noted above, it is often necessary to disrupt a system in order to identify its parts and determine what they contribute to the operation of the whole system. Thus, in a further attempt to isolate specific systems and determine their operations, psychologists sometimes impose additional tasks intended to saturate the processing capacity of other systems and remove them from playing an active role. Going further and experimentally disrupting processing by lesioning sections of the brain is generally prohibited by ethical principles, but an additional valuable source of information in found in naturally occurring brain lesions. Often the precise nature of the lesion is difficult to determine, but even without relying on such information, neuropsychology has been able to use patterns of performance deficits to decompose the cognitive system into distinct processing units. One approach to doing this relies on identifying syndromes, patterns of similar deficits occurring in multiple patients. Another approach relies on either single cases or groups of cases and seeks to discover dissociations between deficits. For example, O'Keefe and Nadel (1978) argue that rats with hippocampal lesions show deficits in one kind of memory system for places while leaving another memory system for places in place. Even more compelling evidence for different processing systems is provided by double dissociations. A double dissociation occurs when, for example, ability *A* is lost in one patient but preserved in another, while ability *B* is lost in a different patient while *A* is preserved. A number of researchers have used information about such double dissociations to show that information about what an object is and where it is are processed independently in visual processing (Kosslyn *et al.* 1990). A number of other double dissociations have been identified in aphasic and dyslexic patients (Shallice, 1988).

How do these efforts to decompose the cognitive system that are already well advanced in cognitive psychology and neuropsychology as well as cognitive neuroscience relate to the development of symbol processing and connectionist models of cognitive performance? The first thing to note is that, as approaches to modeling, connectionist and symbolic approaches are concerned to fit the data about cognitive performance, including data about different processing components that have been isolated. When one examines reports of cognitive simulations in symbolic cognitive science, what one finds are precisely such attempts to evaluate simulation results against data derived from experimental procedures. Connectionist modeling has followed the same pattern. For example, Patterson *et al.* (1989) and Hinton and Shallice (1990) have attempted to account for data regarding processing of written words by normal people and patients suffering various forms of dyslexia.[5]

Given that connectionism and symbol processing represent different frameworks for modeling performance of cognitive systems, we need to explore what are the reasons for adopting one or the other framework. One way to approach this question is to ask where the different modeling approaches draw their inspiration form. As is often the case when a level of inquiry is not fully developed, the models are constructed by analogy from those already in use in related inquires, including those at nearby levels. Examples abound in biology. Biochemistry, as it was developing, built models by analogy with those developed in physical and inorganic chemistry. This was particularly true of the notion of an enzyme, which was modeled on the notion of a physical catalyst. It was a then unproven assumption that enzymes like inorganic catalysts were well-defined molecular structures. The fact that these models were based on analogies and were not already demonstrated to be appropriate is shown by the presence of a competitor. Biocolloidology assumed that the reactions that biochemists were seeking to explain in terms of enzymes were better explained in such colloidal terms as surface tensions. Only subsequent success in identifying, isolating, and purifying enzymes showed that biochemistry, not biocolloidology, would provide the most adequate models. By the time biochemistry achieved a mature form, though, the notion of an enzyme had also been transformed to take into account specific information that had been learned about the macromolecules that constituted enzymes.

To appreciate the relevance of this point to the conflict between connectionism and symbol processing, let us consider the origins of each modeling approach and the reasons for thinking each might offer a plausible account of the structure of cognitive systems. In presenting his case for connectionism, Smolensky differentiated a conscious rule interpreter from an intuitive processor. His notion of a conscious rule interpreter is useful for understanding the conceptual origins of the symbolic approach. Smolensky arrived at his conception of what the conscious rule interpreter did by starting with cultural knowledge. He focused on those features of symbolic representations and human reasoning about them that enable these structures and processes to play an important role in culture:

> This method of formulating knowledge and drawing conclusions has extremely valuable properties:
>
> a. *Public access*: The knowledge is accessible to many people.
>
> b. *Reliability*: Different people (or the same person at different times) can reliably check whether conclusions have been validly reached.
>
> c. *Formality, bootstrapping, universality*: The inferential operations require very little experience with the domain to which the symbols refer (Smolensky, 1988, p. 4).

The last feature is significant in that it allows the system to be employed in totally new domains about which it has acquired little experience. We can, for example, learn something about how to function in a new domain by reading. For our purposes, however, the first two features deserve particular notice. The usefulness of this system depends upon its being public both in its access and in the ability to check on its reliability. This is achieved in large part by allowing the knowledge employed to be represented in *public* symbols, such as the sentences of natural language, and having the use of this information be manifest publicly in terms of actions or the production of new public symbols.

The treatment of the mind as internally processing symbols is very much an adaptation of this public use of symbolic representations and the publicly used rules for evaluating use of this information. In modeling the mind as a symbolic system one takes over the language used for external symbols and the principles governing legitimate inferences based on those symbols and uses this to characterize the internal structure of the mind. There is something very unnatural about this borrowing, however. Typically, in developing a mechanistic explanation, as one takes a system apart, one discovers that the behavior of the parts cannot be described in the same vocabulary as the overall operation. For example, yeast cells perform fermentation, a process described as the production of alcohol from sugar by yeast living in the absence of air. Initially investigators tried to explain fermentation without changing vocabulary: they spoke of the steps in fermentation as themselves fermentations. However, as biochemistry matured it developed its own vocabulary. Biochemists refer to a variety of chemical reactions such as oxidations, reductions, transphosphorylations, etc. and enzymes which catalyze these reactions. Taken independently this vocabulary does not seem to have very much to do with the physiological vocabulary that uses terms such as *fermentation*. It was for this reason that some vitalists could dismiss the relevance of chemistry to physiology. It is the task of the model builder to construct an account that shows how the processes defined in one vocabulary in fact perform processes characterized in the other vocabulary. Thus, it was the various models that were advanced attempting to show how chemical reactions could achieve the overall result of fermentation and other physiological processes that connected the two vocabularies.

The general point here is that causal processes at different levels in nature are generally quite different in character and one must develop appropriate vocabularies to describe the particular causal interactions at any given levels. Using the symbol processing framework to model the internal operations of a cognitive system, however, violates this principle. In this case, the language developed and appropriate for characterizing processes at a higher level may also turn out to be the correct language for characterizing processes at a lower level, but this would be very unusual. On the other hand, since representations and rules are items with which we are already familiar from being systems that use them in our cultural practices, this would be a natural starting point in developing cognitive models. Moreover, if the representational systems we learn (such as natural language) are partly characterized by rules about the construction of those symbols (e.g., rules about how to form plurals and past tenses of words), then it is not surprising that by using rule systems we are able to develop models that simulate our use of these representational systems. This success, however, does not insure that the rule system captures the structure of the processing system that enables us to use these representational systems. So, while symbolic models are a natural starting point, and may simulate quite accurately the behavior, the doubts raised above about a lower level system employing the same architecture as a higher level system should give one pause.

Connectionism draws its inspiration from a different source. The units and connections posited in connectionist models were inspired by early analyses of brain organization. Since neuronal systems are clearly systems within cognitive agents, basing cognitive models on neuronal systems may seem more justified than basing them on higher level processes such as conscious rule interpretation. The adoption of a neural model would be quite justified if it were thought that the neural level was the correct level at which to model cognitive processes. However, most connectionist who are interested in cognitive modeling do not accept the idea that units in connectionist simulations should be equated with neurons. Smolensky, for example, rejects the view that connectionism advances a theory at the neural level, pointing to a number of substantive disanalogies between neural systems and connectionist systems. Among the differences he notes are the following: neural systems seem to employ a dense pattern of connectivity to nearby neurons and a highly specific mapping between more distant neurons, while connectionist nets employ a uniform pattern of connectivity; neural systems use multiple signal types, whereas connectionist systems use a single signal type; and neural systems employ an intricate procedure of signal integration at individual neurons while connectionist systems use a linear procedure (Smolensky, 1988, p. 9). He also notes, quite correctly, that many advances being made in connectionism are the result of attempting to account for cognitive phenomena, not the result of attempts to provide more realistic neural models. For example, the introduction of modularity into networks (Nowlan, 1990, Jacobs *et al.*, 1991) is partly the result of attempts to overcome catastrophic interference (McCloskey and Cohen, 1989). Thus, the overriding goal in connectionist modeling is not to characterize neural systems, but to develop a framework useful for describing cognitive performance. Most connectionist seem implicitly to assume that connectionist modeling describes activities at a higher level than actual neural systems. Individual units are sometimes portrayed as performing the function carried out by ensembles of actual neurons. But then to appeal to a neural-like architecture clearly assumes that at a level above actual neural systems we will find structures with the same architecture as neural systems (i.e., systems consisting of simple processing units and activations passed between them). But it is not clear why we should expect higher level structures and operations to be sufficient similar to neural processes so as to be well characterized in terms of networks of simple processing units as connectionism proposes. A model based on analogy of the cognitive processing system to neural systems seems no more likely to be true than one based on analogy to the conscious rule interpreter.

Even if one doubts that connectionist networks characterize the structure of the cognitive system and holds that they are better viewed as providing abstract accounts of neural processing, one might at least hold out hope that connectionist modeling might provide a useful strategy for discovering the appropriate higher level structures and processes. Increasingly, work in connectionism has turned away from simple networks to much more structured, modular networks (for a review of some of these, see Bechtel, 1993a). It may be that as further research points to more elaborate structures in networks that we might discover the sorts of higher level structures that do figure in explaining cognitive phenomena. Moreover, it has sometimes been claimed by connectionists that the power of the neural analogy is to open up a space of alternatives to symbolic models. Connectionist systems that have been developed to date represent just a small percentage of the possible range of dynamic systems, and provide us an entree to this range within which we might find more suitable systems for modeling cognition.

This strikes me as at least a plausible scenario. But discovering such higher level structures may not be a straightforward process. The analogy to research on oxidative phosphorylation noted above may indicate the difficulties that lie ahead. It now appears that some variant on Mitchell's (1966) chemisomotic hypothesis is close to correct according to which membranes play an important role in the process by creating ion gradients which then drive the phosphorylation of ADP to make ATP. The important point here is that membranes are structures at a higher level than the structures traditionally considered in biochemical models. These membranes are physical structures built from chemical components, but their significance for explaining oxidative phosphorylation was not discovered by traditional biochemical theorizing, which focused on identifying chemical intermediates. The alternative idea of developing an ion gradient and using that to build ATP was motivated by analogy to models in fields other than traditional biochemistry and imported in. If cognitive studies follow a similar pattern, then building up from neuron-like systems as found in contemporary connectionist models will not alone point to the kinds of structures and processes needed to model cognition.

What these considerations seem to imply is that neither connectionist nor symbolic systems are likely to provide the appropriate frameworks for modeling cognitive processes. Each is based on a framework that is appropriate to a different level of theorizing, symbolic systems for the level at which humans function as conscious rule interpreters and connectionist systems for the level of neural processing. If there is in fact a level of cognitive processing that occurs within the person but above the level of neural processes, then a different set of concepts and modeling tools is needed to develop these models. There is a level

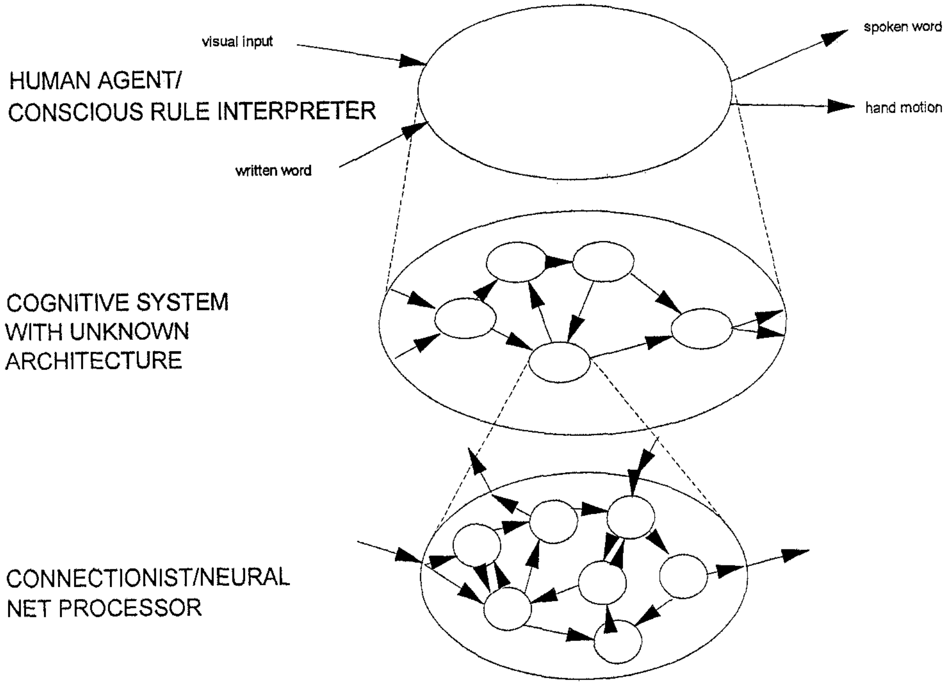

Fig. 5. The missing level: The cognitive system located between the human agent/conscious rule interpreter and the connectionist/neural processor.

of organization for which we do not yet know the nature of the basic causal interactions and for which we have yet to develop an adequate vocabulary (Figure 5). What is needed is further creative theorizing and empirical investigation to identify the character of the cognitive processes which are realized in a neural system and which explain the behavior of cognitive agents, including their activities as conscious rule interpreters.

5. Conclusion

One difficulty we face in answering the question as to what are the relevant levels of organization responsible for producing cognitive phenomena and determining how they relate to other levels is that we are still at a very early juncture in developing empirical theories of cognitive phenomena. We have, at present, only a few tools for decomposing experimentally the cognitive system to determine what its parts are and how they contribute to the performance of cognitive tasks. Identifying levels in nature, characterizing the processes at these levels, and building models of how the components identified at a level perform a higher level task, are not activities that can be accomplished by philosophical inquiry. All I can hope to have done here is to clarify the nature of that task by analyzing how levels do figure in scientific inquiries and showing how current endeavors could be envisaged to fit into the sorts of multi-level inquiries common in science.

My discussion here has identified a manner in which connectionism can be seen to occupy a lower level in nature than symbolic theorizing. Symbolic theorizing concentrates on external symbols and ways humans as agents manipulate those symbols. If we pursue the strategy, about which I have raised doubts, of using symbol processing accounts appropriate for such public use of symbols to try to account for inner cognitive operations, then it becomes far less clear that connectionism and symbol process accounts are at different levels. They appear more as competitors. But even if we restrict symbolic models to the level of public use of symbols, it remains unclear where to place connectionism. Insofar as connectionist models are built upon analogy with neural processing, it might be best to keep them at that level, and construe them perhaps as offering fairly abstract accounts of such processing. But then it is not clear that they are appropriate for modeling cognitive phenomena. Cognitive models likely involve structures above the level of neural processes and there is no reason to think that the structural architecture at that level will be very similar to that at the neural level. The characteristics of such a level would have to be discovered. Of course it may turn out that no extra level is needed. In that case cognitive inquiry and neural inquiry would collapse and connectionist or more sophisticated network models would suffice for describing the nervous system and explaining how it performs cognitive activities.

Notes

[1] An earlier version of this paper was presented as part of the Fifteenth Annual Greensboro Symposium in Philosophy (April 1991) and to the Cognition Project at Emory University. I thank members of both audiences, anonymous referees for this journal, and especially Adele Abrahamsen for useful comments and suggestions.

[2] Of these terms, *connectionism* is the more generic, referring to any of the class of cognitive models that involve nodes that acquire activations and connections which transmit these activations to other nodes. The terms *parallel distributed processing* is usually reserved for the approach to connectionism developed by David Rumelhart, Jay McClelland and their colleagues in which nodes do not individually serve a representational function, but in which representations are patterns of activations across collections of nodes.

[3] This sentence actually introduces a caveat: the connectionist account and the symbolic account might not be about the same thing. This is in fact the diagnosis Rosenberg offers, using Smolensky's own distinction between a conscious rule interpreter and an intuitive processor. For Rosenberg, the symbolic paradigm is prototypically concerned with explaining how humans reason with concepts, which Rosenberg takes as the proper delineation of cognition. The subsymbolic paradigm is concerned with how humans perform a variety of intuitive tasks, such as those accomplished by individuals who possess expertise in a given domain. Rosenberg's solution may be less than a happy one, however, for the very reason that Smolensky avoided what he termed the *cohabitation* approach. It not only seems problematic to assume that the mind would have developed a totally different system to handle conscious conceptual reasoning, but it also would seem necessary that the mind have ways of relating

results of conceptual reasoning to the sorts of intuitive reasoning that connectionism is designed to explain.

[4] Fodor, who does view connectionism as potentially an account of implementation, however, should not accept this view. He has long objected to the reductionist scenario on the grounds that there will not be bridge laws relating predicates of the special sciences and those of the more basic sciences (see Fodor, 1974).

[5] Not all connectionist simulations attempt to model experimentally differentiated psychological functions. Many connectionists, as many researchers in other areas of AI, have relied on very intuitive conceptions of what might constitute distinctive cognitive performances. Thus, Rumelhart and McClelland (1986) modeled the formation of the past tense in English as if it were a totally separate task from other components of language processing. As Pinker and Prince (1988) argue, however, this task is likely to be closely integrated with other language processing tasks. This is not to say that McClelland and Rumelhart are not concerned to capture the data generated by human performance. Indeed, they defend their model in terms of its capacity to account for data about human acquisition of the English past-tense. Pinker and Prince, in challenging the model, present other data they claim it cannot account for. Such is a common form of interaction over proposed models, both in cognitive science and in other fields.

References

Barlow, H.B. (1953), 'Summation and Inhibition in the Frog's Retina', *Journal of Physiology* 119, pp. 69–88.

Bechtel, W. (1988), *Philosophy of Science: An Overview for Cognitive Science*, Hillsdale, NJ: Lawrence Erlbaum Associates.

Bechtel, W. (1993a), 'Currents in Connectionism', *Minds and Machines* **3**, pp 125–153.

Bechtel, W. (1993b), 'Integrating Sciences by Creating New Disciplines: The Case of Cell Biology', *Biology and Philosophy* **8**, pp. 277–299.

Bechtel, W. and Abrahamsen, A.A. (1991), *Connectionism and the Mind*, Oxford: Basil Blackwell.

Bechtel, W. and Richardson, R.C. (1992), *Discovering Complexity: Decomposition and Localization as Scientific Research Strategies*, Princeton, NJ: Princton University Press.

Broadbent, D. (1985), 'A Question of Levels: Comment of McClelland and Rumelhart', *Journal of Experimental Psychology: General* **114**, pp. 189–192.

Causey, R. (1977), *Unity of Science*, Dordrecht: Reidel.

Darden, L. and Maull, N. (1977), 'Interfield Theories', *Philosophy of Science* **43**, pp. 44–64.

Fodor, J.A. (1974), 'Special Sciences (Or: Disunity of Science as a Working Hypothesis', *Synthese* **28**, pp. 97–115.

Fodor, J.A. and Pylyshyn, Z.W. (1988), 'Connectionism and Cognitive Architecture: A Critical Analysis', *Cognition* **28**, pp. 3–71.

Hinton, G.E. (1986), 'Learning Distributed Representations of Concepts', *Proceedings of the Eighth Annual Conference of the Cognitive Science Society*, Hillsdale, NJ: Lawrence Erlbaum, pp. 161–187.

Hinton, G.E. and Shallice, T. (1991), 'Lesioning an Attractor Network: Investigations of Acquired Dyslexia', *Psychological Review* **98**, pp. 74–95.

Hubel, D.H. and Wiesel, T.N. (1962), 'Receptive Fields, Binocular Interaction and Functional Architecture in the Cat's Visual Cortex', *Journal of Physiology* **166**, 105–54.

Hubel, D.H. and Wiesel, T.N. (1968), 'Receptive Fields and Functional Architecture of Monkey Striate Cortex', *Journal of Physiology* **195**, pp. 215–243.

Jacobs, R.A., Jordan, M.I., and Barto, A.G. (1991), 'Task Decomposition Through Competition in a Modular Connectionist Architecture: The What and Where Vision Tasks', *Cognitive Science* **15**, pp. 219–250.

Kosslyn, S.A., Flynn, R.A., Amsterdam, J.B., and Wang, G. (1990), 'Components of High-Level Vision: A Cognitive Neuroscience Analysis and Accounts of Neurological Syndromes', *Cognition* **34**, pp. 203–277.

McCauley, R.N. (1986), 'Intertheoretic Relations and the Future of Psychology, *Philosophy of Science*, **53**, pp. 179–199.

McClamrock, R. (1991) 'Marr's Three Levels: A Re-Evolution', *Minds and Machines* **1**, pp. 185–196.

McClelland, J.L., Rumelhart, D.E., and the PDP Research Group (1986), *Parallel Distributed Processing: Explorations in the Microstructure of Cognition. Vol. 2: Psychological and Biological Models*, Cambridge, MA: MIT Press.

McCloskey, M. and Cohen, N.J. (1989), 'Catastrophic Interference in Connectionist Networks: The Sequential Learning Problem', in G.H. Bower (ed.) *The Psychology of Learning and Motivation*, vol. 24, New York: Academic Press, pp. 109–65.

Marr, D. (1969), 'A Theory of Cerebellar Cortex', *Journal of Physiology* **202**, pp. 437–470.

Marr, D. (1982), *Vision. A Computational Investigation into the Human Representation and Processing of Visual Information*, San Francisco: W.H. Freeman.

Mitchell, P. (1966), 'Chemiosmotic Coupling in Oxidative and Photosynthetic Phosphorylation', *Biological Reviews* **41**, pp. 445–502.

Nagel, E. (1961), *The Structure of Science*, New York: Harcourt, Brace.

Nickels, T. (1973), 'Two Concepts of Intertheoretic Reduction', *Journal of Philosophy* **70**, pp. 181–210.

Nowlan, S.J. (1990), 'Competing Experts: An Experimental Investigation of Associative Mixture Models', Technical Report CRG-TR-90-5, Department of Computer Science, University of Toronto.

O'Keefe, J. and Nadel, L. (1978), *The Hippocampus as a Cognitive Map*, Oxford: Clarendon Press.

Pinker, S. and Prince, A. (1988), 'On Language and Connectionism: Analysis of a Parallel Distributed Model of Language Acquisition', *Cognition* **28**, 73–193.

Rosenberg, J.F. (1990). 'Treating Connectionism Properly: Reflections on Smolensky', *Psychological Research* **4**, pp. 163–174.

Rumelhart, D.E. and McClelland, J.L. (1985), 'Levels Indeed! A Response to Broadbent', *Journal of Experimental Psychology: General* **114**, pp. 193–197.

Rumelhart, D.E. and McClelland, J.L. (1986). 'On Learning the Past Tense of English Verbs', in J.L. McClelland, D.E. Rumelhart, and the PDP Research Group (eds.) *Parallel Distributed Processing: Explorations in the Microstructure of Cognition. Vol. 2: Psychological and Biological Models*, Cambridge, MA: MIT Press.

Rumelhart, D.E., McClelland, J.L., and the PDP Research Group (1986), *Parallel Distributed Processing: Explorations in the Microstructure of Cogntiion. Vol. 1: Foundations*, Cambridge, MA: MIT Press.

Schaffner, K. (1967), 'Approaches to reduction', *Philosophy of Science* **34**, pp. 137–47.

Shallice, T. (1988), *From Neuropsychology to Mental Structure*, Cambridge, England: Cambridge University Press.

Smolensky, P. (1988). 'On the Proper Treatment of Connectionism', *Behavioral and Brain Sciences* **11**, pp. 1–23.

Wimsatt, W.C. (1976), 'Reductionism, Levels of Organization, and the Mind-Body Problem', in G. Globus, G. Maxwell, and I. Savodnik (Eds.), *Consciousness and the Brain: A Scientific and Philosophical Inquiry*, New York: Plenum, pp. 205–267.