

## INTERTHEORY RELATIONS IN COGNITIVE SCIENCE: PRIVILEGED LEVELS AND REDUCTIVE STRATEGIES

JESÚS EZQUERRO

Departamento de Lógica y Filosofía de la Ciencia  
Universidad del País Vasco  
ylpezmaj@sf.ehu.es

FERNANDO MARTÍNEZ MANRIQUE

Departamento de Filosofía  
Universidad de Granada  
fernan\_martinez@yahoo.com

**SUMMARY:** Research in cognitive science has often assumed the existence of a privileged level that unifies theoretical explanations arising from different disciplines. Philosophical accounts differ about the locus of those intertheory relations. In this paper, four different views are analyzed: classical, connectionist, pragmatist, and reductionist, as exemplified in the works of von Eckardt, Horgan and Tienson, Hardcastle, and Bickle, respectively. Their divergences are characterized in terms of the possibility of such a privileged level. The classical view favors a privileged computational level. The connectionist revision tries to draw biological intuitions into computational models to provide a link between neurons and computations. The pragmatist approach rejects the idea of a privileged level and offers a more eclectic view of cognitive theory building. The reductionist account pursues theoretical unification by means of the reduction of higher level theories to basic level (biological) ones.

**KEY WORDS:** cognitive science, computationalism, connectionism, reduction

**RESUMEN:** La investigación en ciencia cognitiva suele asumir la existencia de un nivel privilegiado que unifica las explicaciones teóricas provenientes de distintas disciplinas. Las interpretaciones filosóficas difieren respecto al locus de tales relaciones interteóricas. En este artículo se analizan cuatro concepciones: clásica, conexionista, pragmática y reduccionista, ejemplificadas respectivamente en las obras de von Eckardt, Horgan y Tienson, Hardcastle y Bickle. Sus divergencias se caracterizan en términos de la posible existencia del nivel privilegiado. El enfoque clásico propugna un nivel privilegiado computacional. La revisión conexionista importa intuiciones de la biología a los modelos computacionales, con la intención de vincular lo neuronal con lo computacional. El enfoque pragmático rechaza la idea de un nivel privilegiado y ofrece una visión más ecléctica de la teorización cognitiva. La concepción reduccionista promueve la unificación teórica mediante la reducción de teorías de alto nivel a teorías de nivel básico (biológico).

**PALABRAS CLAVE:** ciencia cognitiva, computacionalismo, conexionismo, reducción

### 1. *Introduction*

Perhaps the most distinctive feature of cognitive science is its *interdisciplinary* nature. This is not a well-defined notion. On the one hand, it involves more than acknowledging that there are sciences which study related phenomena, and that results from one of them can illuminate the research carried out in another. On the other, it is difficult to specify what kind of relation the different sciences sustain so that we can talk about genuine *interdisciplinarity* instead of a loose *multidisciplinarity*. The latter implies just a (more or less lax) relation between disciplines that pursue their own theoretical goals, and it is canonically expressed in Gardner's famous hexagon (Gardner 1985).<sup>1</sup> Genuine *interdisciplinarity*, by contrast, suggests that the relation extends *into theories themselves*, so that, in order to achieve their explanatory task, theories would go across several disciplines. There has been in cognitive science a vacillation between both senses.<sup>2</sup> Moreover, it is not even clear what kind of normative status should be assigned to *interdisciplinarity*. If we take a strong normative stand, we might say that it is a *necessary* property of cognitive theories, i.e., that we ought to regard as cognitive only those theories that comprise several levels of theorizing, linked by such and such relations. If we relax our standards, we could consider *interdisciplinarity* as a sort of regulative idea, a vague desideratum that would be *convenient* to meet. Or we may take a thoroughly descriptive stand and insist that, rather than a property of theories, *interdisciplinarity* is simply an unfortunate *consequence* of the current meager development of cognitive science, a consequence that would disappear in future solid scientific progress. Finally, we may want to argue that *interdisciplinarity* is neither a property that theories currently have nor one that they ought to possess in the future,

<sup>1</sup> Gardner's six disciplines are philosophy, psychology, artificial intelligence, linguistics, anthropology, and neuroscience.

<sup>2</sup> Cf. Hardcastle 1996, Bechtel 1988 and 1994, or Bechtel and Richardson 1993. Other authors have pointed out a similar contrast between cognitive science and cognitive sciences, e.g., Rivière 1991.

but only a philosopher's *phantasm* when she daydreams about unifying everything from quantum physics to sociological data.

A cursory look at cognitive research allows us to discard the last gloomy judgment, at least with respect to some areas of research. Thus, we can find disciplines such as neuropsychology or neurolinguistics that pertain to more than one level of theorizing: the former studies the relation between brain states and behavior, while the latter tries to establish links between language, on the one side, and brain function and cognitive capacities, on the other. Correlations between syntactic processing deficits and certain brain malfunctions, such as Broca aphasia, are examples of the kinds of correlations that those theories look for. In the process of formulating their hypotheses they find themselves drawing elements from different vocabularies and postulating connections between them.

We could call these correlations "vertical", since they relate elements that are on different levels. But there are also "horizontal" relations that generate theories midway between different disciplines. This is what Darden and Maull (1977) call *interfield* theories, or what Abrahamsen (1987) characterizes in terms of "bridging boundaries" as opposed to "breaking boundaries". In this case, one discipline is reconceptualized in terms of another (e.g., Chomsky's attempt to psychologize linguistics), while in interfield theories there are connections that can be useful for several disciplines (e.g., categorization theories like those from Rosch (1978), Medin and Schaffer (1978) or Lakoff (1987) are located in the middle of psychology, linguistics, and anthropology).

Another difficulty is how to account for the interdisciplinary relations between theories. Should they be regarded as a reduction of a discipline to a more basic one? Or is there a different kind of relation more amenable to the pluralism that cognitive studies often endorse? If we consider the case of breaking boundaries, for instance, it seems quite natural to treat it in terms of reductive relations, but this is not so clear with respect to bridging boundaries. Here we do not find that a field's ontology or explanatory resources are mapped onto a different

one. Rather, what we have are extensions<sup>3</sup> of the theories so that their scope gets enlarged. Yet, results from theories that bridge boundaries are usually taken as the basis to construct theories of mental architecture, so we could say that they enter into reductive relations, even if indirectly, inasmuch as architectural pillars are formulated reductively. What is needed in these cases is to explain phenomena obtained at one description level, e.g., anthropological data used by folk psychology, in terms of theories about the internal structure of mental processing, theories which may postulate strong hypotheses about, say, how concepts are stored in memory.<sup>4</sup>

The above difficulties increase when we add the ones put forward by the Philosophy of Mind along the years. To begin with, it is well known that the classical account of intertheoretical relations, Nagel's model of reduction, was intended as an extension of Hempel's covering law model of explanation of events, in order to explain laws and theories. Following this account, theory reduction consists in deducing the laws of the candidate theory to be reduced from the laws of the reducing theory. Given that both the reducing and the to-be reduced theories usually have different vocabularies, in order to allow deduction, the employment of bridge laws connecting their terms becomes necessary. However, it turns out that some philosophers have questioned the existence of proper laws in psychology, or the possibility of deducing them from more basic laws, for example, from the laws of brain sciences, or both. The main problem is the very possibility of having available bridge laws connecting properties at the psychological level and properties at the brain sciences level.

<sup>3</sup> The most interesting extensions are explanatory extensions (Kitcher 1981), which we will consider later.

<sup>4</sup> It is well worth noting, in this respect, that Rosch (1978) openly opposed the extrapolation of her results (an account of prototypicality obtained from the analysis of categorization of colors) to hypotheses about processing structures and memory storage. However, it is well known that many psychologists used them with this purpose. Indeed, connectionist modeling takes such an assumption.

There are at least two well-known arguments against the possibility of lawful connections between psychological and brain levels. The first one comes from Donald Davidson's defense of his anomalous monism. Davidson argued at length that psychological and physical predicates cannot be put together. The language game of psychological predicates presupposes holism and rationality. The consequence of these essential features is that psychological laws are *soft*, or *ceteris paribus* laws. Psychophysical laws would have, therefore, the same condition, given that they make essential use of psychological predicates. Physical laws, on the contrary, are strict. For Davidson, a 'strict law' is one which makes no use of open-ended escape clauses such as "other things being equal". Davidson's claim is that without strict psychophysical (bridge) laws, there is no possibility of reducing psychology at all.<sup>5</sup>

The second argument comes from Jerry Fodor (1974). This argument assumes that psychological predicates must be functionally defined. This assumption is widely accepted in the cognitive science community. The problem, Fodor claims, is that functional predicates can be instantiated by multiple and heterogeneous physical realizers. Therefore, if we were to reduce a psychological law, the result would be a conditional sentence composed by a disjunction of heterogeneous physical predicates (properties), both in the antecedent and in the consequent. Fodor thinks that a disjunctive law of this sort cannot be a scientific law. In a nutshell, Fodor's claim is that psychological kinds are *natural kinds*, and thus the laws connecting them are good laws, although they are *ceteris paribus*. But we should set aside the hope of reducing them, given the lack of available psychophysical laws.

<sup>5</sup> See Davidson 1970, 1973, and 1974. Indeed, Davidson has repeatedly claimed that there is no possibility of psychophysical (bridge) laws. But it is easy to see that the only valid conclusion of his argument is that psychophysical laws would be soft, *ceteris paribus* laws. Lately, Davidson (1993) accepted that point. Furthermore, carefully examined, his argument can be generalized. Since his claim is that only basic science has strict laws, and given that all cases of reduction involve two theories, one of which must be, necessarily, non-basic, any attempt of reducing would be impossible.

The above picture gets more complicated when it is translated into modern cognitive science. As we will see in the next section, the assumption that cognition is information processing brought about a specific way of conceiving psychological theories. This way is the computational approach. According to David Marr (1982), we can distinguish three levels in every computational system: the computational, the algorithmic and the implementation levels. But it turns out that each level is multiply realizable by its immediate lower level. Therefore, the problem of multiple realizability appears again. In so far as Marr's three level partition can be applied to every computational system, we can apply it to connectionist systems as well. In this concern, some defenders of connectionism have contended that connectionist algorithmic level is a fine grained realization of coarse classical computational states. Classicists, on their part, have proposed that connectionism can merely serve as the physical implementation level of classical architecture.<sup>6</sup> This issue is still under dispute.

In short, the landscape that the development of cognitive science offers to the philosopher of science contains intertheory relations which are much more complex than those usually considered in the literature. The most notable elements that we find in the new situation are the following ones: (a) scarcity of laws; (b) abundance of functional properties; (c) multiple realization; (d) the specific problem of relating levels in Marr's sense; (e) reductions that involve ontological elimination together with non-eliminative reductions; (f) the problem of how to relate interfield theories both in horizontal and vertical ways; and (g) the problem of evaluating the possibility of unified cognitive architectures that integrate the relevant factors.

All these problems seem to pose a heavy burden for the accounts of intertheory relations and of reduction that are usually adopted in philosophy of science. In this paper we will analyze

<sup>6</sup> See Smolensky, Legendre and Miyata 1993 on the connectionist side, and Fodor and Pylyshyn 1988 on the classicist one. The position defended by Smolensky *et al.* could be interpreted as suggesting something like a micro-macro relation of reduction between classicist and connectionist laws, but there is no place in this paper to analyze this possibility.

four different conceptions about the structure of cognitive science. We will label them as the *classical* view, the *connectionist* revision, the *pragmatist* approach, and the *reductionist* account. We will examine a paradigmatic version for each of them, versions that correspond, respectively, to the works by Barbara von Eckardt (1993), Terence Horgan and John Tienson (1996), Valerie Hardcastle (1996), and John Bickle (1998). One of the issues in which they differ is the *locus* of intertheory relations, that is, the point at which the different theories connect so that we can talk of a single unified account. In this respect, we can characterize the divergence in terms of the question of whether we ought to privilege a level to build our cognitive theories.

## 2. *The Notion of Privileged Level*

Research in cognitive science has often assumed the existence of a *privileged level* at which all the different disciplines come to converge. Computational theories were the first ones to offer themselves as such a level. The equation “cognition = information processing”, on which the cognitive revolution was founded, seemed to find its natural place in the technical and mathematical developments provided by those theories. The possibility of obtaining a system of computational mechanisms that accounted for the totality of cognitive phenomena offered the promise of a “unified theory of cognition” (Newell 1990). On the other hand, Marr’s (1982) elegant distinction of three levels was nicely suited for the computational paradigm. Marr distinguished between (1) a computational level, (2) an algorithmic level, and (3) an implementational level. Level 1 provides an analysis of the task to be done, i.e., the “what” of the computation. Level (2) specifies the algorithms in which the computation was to be performed, i.e., the “how” of the computation. Level (3) details the particular mechanism that supported the specified algorithm, so to speak, the “how and where” of the computation.<sup>7</sup>

<sup>7</sup> Marr’s distinction received soon a number of qualifications and terminological variations, trying to accommodate other intuitions. Newell (1982, 1990) identifies a knowledge level, a symbolic level and a physical level, with

Even if Marr considered that the three levels needed to be worked out in the process of solving certain computational problems, many researchers soon assumed that it was in level 2 where the most interesting issues arose. In particular, it was at this level where algorithmic accounts could be tied to representational questions. For instance, we can consider the problem of recognizing the written shape of a letter (the “what”), a problem that can be solved in several ways (the “how”). Each different way to solve the problem can be identified with an algorithm, and the elements involved in the transitions of the algorithm can be regarded as representations of the letters. So each algorithmic specification offers at the same time a different representational partition (e.g., different possible segmentations of the letters in vertical lines, horizontal lines, etc). Finally, each algorithm can be implemented in more than one way, depending on the characteristics of the mechanism chosen to that end.

However, there were other candidates to provide the privileged level of cognitive theorizing. The most relevant ones are neurosciences and cognitive psychology (without computational clothing). The former can claim that unification is guaranteed by its more basic position in the hierarchy of levels, while the latter can resort to its natural capability to capture what is more specific of mental life, that is, its intentionality. It is not surprising that each alternative can be located at one of the other two levels of Marr’s schema. Postulating neurosciences as the privileged level is consistent with a reductive tendency: the idea is that we can establish vertical links between the elements and relations that appear in upper levels and the elements and relations that appear in a biological description. In

a hierarchy of sublevels running from bottom to top (based on computer notions, such as register transfer, logical circuit, etc). Pylyshyn (1984) distinguished the semantic, syntactic (symbolic) and mechanism levels. Dennett (1987) introduced the idea of intentional level to refer to level 1. Sterelny (1990) reserves the term ‘computational’ for level 2, while labeling level 1 as “ecological” (to encompass the descriptions of external behavior). Corbí (1993) distinguishes between two description levels in folk psychology, a level of mechanism that can be classical or connectionist, and a neurological level. Horgan and Tienson (1996) extend the three levels to accommodate mathematical possibilities other than algorithms (we will say more on this later).

contrast, favoring psychology usually involves a thoroughly anti-reductionist lean, which maintains the autonomy of intentional phenomena with respect to the underlying substratum. Part of the initial appeal of computational approaches was that they seemed to provide a middle way between both these extremes. In this view psychology could be non-reductionist while still constrained by the architecture of the system that implemented it. In fact, computational psychology lost part of its initial momentum when problems about the relation between levels began to emerge.<sup>8</sup>

A further problem with Marr's scheme is that it can be regarded as a general framework to formulate intertheory relations in cognitive science, only if we could assume that each level corresponds to one of the disciplines whose theories we want to relate. Yet, Marr conceived of his distinction as a tool of analysis *within* each discipline. More specifically, he considered it a tool for computational approaches themselves. So the assumption that level 2 and computational psychology are the same thing is an unwarranted simplification, sustained by the confusion between an epistemological and an ontological sense of level.<sup>9</sup> From an epistemological point of view, a level is just a particular way of looking at nature, that depends on the interest, tools, etc., that constrain our looking. It is a convenient conceptualization *for* studying any phenomenon. Hence, if we find it useful, we may use Marr's three levels to analyze a neural process. The ontological sense, however, assumes that a level is a particular way in which nature is organized. As the organization becomes more complex, new entities and properties appear, and new vocabulary and theoretical resources are needed to account for

<sup>8</sup> We can see the problem of psychosemantics in these terms (Fodor 1987), which is basically the problem of relating the intentional and the physical. It is not enough to resort to the notion of symbol as an intermediate, since what is in question is the relation between the symbol and the symbolized. Another problem of levels is how to flesh out the notion of implementation, to account for the relation between levels 2 and 3 (Foster 1992).

<sup>9</sup> A discussion about different notions of levels (structural, functional, ontological, etc.) can be found in Bechtel 1994, Butler 1994, and Rosenberg 1994.

them. So it can be said that each different level demands a different discipline. Hence, if a neural process occurs at level 3, we cannot analyze it in level 2 terms. Even if we may make initial assumptions (constrained also by our epistemological resources) about what levels are really there, the final answer is something to be empirically discovered *in the course of* studying a phenomenon. Thus, Marr's schema may or may not capture the layers of organization of cognitive processes, but if it does not, then it is a failed attempt. As we will see, both senses of level often appear intertwined.

The idea itself of a privileged level is equivocal. Is it a “no-man's land” that provides a common vocabulary? Or is it one of the particular sciences in whose concepts we must regiment the results obtained in the other sciences? Both interpretations are problematic. The former seems to respect the pluralism of an interdisciplinary project, but it is unclear whether there is room for the common terrain that it propounds. The latter finds this terrain in one of the established sciences at the risk of the autonomy of the others, since everything that was not translatable to the “basic science” would be out of cognitive theorizing. The four positions we are going to analyze exhibit different attitudes toward the idea of a privileged level. The classical view, as one can expect, favors the idea of a privileged computational level. The connectionist revision, still within a broad computational stance, tries to draw biological intuitions into artificial models so as to provide a link between neurons and computations. The pragmatist approach rejects the idea of a privileged level in order to offer a more eclectic view of cognitive theory building. The reductionist account, finally, pursues theoretical unification by means of the reduction of higher level theories to basic level (i.e., biological) ones.

### 3. *The Classical View*

As we said above, the notion of computation gave the cognitive revolution the promise of a common ground on which all cognitive phenomena could be described and explained. The rehabilitation of mentalism against behaviorist excesses had also put the

idea of representation into circulation. Both notions came hand in hand. The distinguishing property of mental life is intentionality, i.e., its “aboutness”, its possibility to refer to another part of reality. The key idea, thus, was that some internal physical structures could have a “double life” connected, on the one hand, to other internal structures and, on the other one, to the external world. Structures like these would be representations, and the processes that govern the transitions from some structures to others would be computational operations that govern the transformation of a expression into another. Together, both elements could provide the basis to construct a picture of mental life. To explain a particular mental phenomenon, say, how we remember words, how we categorize objects and so on, required giving the right computational account between the right kinds of representations, which were conceived as symbol systems. So the *physical symbol system* hypothesis (Newell and Simon 1976; Newell 1980) was for a time, explicitly or tacitly, the basis for a general theory of (artificial and human) intelligence and, hopefully, of a scientific theory of mind. Yet, the idea was very much dependent on the state of the art in computer science. In particular, it was closely linked to von Neumann architectures. The emergence of new computational systems, such as artificial neural networks, the dissatisfaction with the traditional computational approach when some of its weaknesses began to be manifest, and the increasing vigor of neuroscientific approaches to cognition, forced scientists to rethink the key concepts of computation and representation. In this section we will examine a conservative reconstruction of cognitive science that tries to offer a broad classical framework capable of accommodating the new elements. The following section will deal with an alternative reconstruction much more revisionist in spirit.

Reconstructing a framework of shared commitments for cognitive science is the main task of von Eckardt (1993). To this end she takes as the paradigms those clear cases on which there is substantial agreement about their place as prominent examples of cognitive science research. One distinctive feature of cognitive science, she claims, is that it is an immature field.

This is not only due to the fact of its relative youth, but also to the fact that it is still “rooted in our ordinary conception of the world”. In other words, folk psychological factors play an important role in shaping theories. The path to maturity will be paved by conceptual changes and technological advances. Is it possible to analyze a science that is in continuous flux? It is, von Eckardt thinks, if we render the analysis in terms of a *research framework*. This is a concept that allows us to reveal the structure of a given science at a particular moment. What it demands is that we can fill the blanks in four foundational elements: domain-specifying assumptions, basic empirical research questions, substantive assumptions which constrain answers to these questions, and methodological assumptions. So the immaturity of the field is not a hindrance for the analysis, which can be used indistinctly for mature sciences.<sup>10</sup> A second distinctive feature of cognitive science is its interdisciplinarity. The research framework will have to be common to all cognitive scientists regardless of their primary disciplinary home. But practice does not necessarily conform to this idea. So, rather than a descriptive reconstruction, what von Eckardt has in mind is a *rational* reconstruction.

For the purposes of the analysis of intertheory relations we will concentrate on substantive assumptions, because it is here where we can find the elements that, allegedly, link cognitive explanations. There are two substantive assumptions: the computational and the representational assumptions. Accordingly, to relate two theories will require us to find a common computational-representational account for them. We will further restrict our focus to the computational assumption, since there are convincing arguments that, without a theory of psychosemantics, the vertical links between scientific psychology (and lower levels) and folk psychology cannot even begin to have an explanation. In other words, without a naturalistic theory about how objects can have intentional properties it is very

<sup>10</sup> Silvers (1996) criticized the criteria used by von Eckardt to distinguish mature from immature science comparing them with the account of research tradition offered by Laudan (1977).

unlikely that intentional descriptions can ever be made theoretically consistent with information-processing descriptions. So von Eckardt must be on the right track when she gives the mental representation assumption a central position in cognitive theorizing, even if she herself does not develop a theory of representation. The computational assumption, however, raises more problems, as we will show.

Von Eckardt affirms that computationalism is a metaphor: “cognitive scientists assume not that mind/brains are literally computers but only that they have some properties in common with them” (p. 104). The computational metaphor is not just loose talk: it is a theory-constitutive analogy. It introduces new theoretical terminology that will constitute the basis for formulating hypotheses and models about the phenomena we want to study. The notion of computer she has in mind is not the abstract notion characterized by computability theory, but the actual devices used by computer scientists. This makes her proposal relative to the kinds of devices available at the moment of analyzing the research framework. At the time she wrote her book, there are only two main competitors in the computational arena: conventional and connectionist machines. So any account of computation, she argues, will have to be broad enough to encompass both systems. There is an obvious risk in this sort of reasoning, since substantive assumptions come to be confused simply with the working hypotheses assumed temporally in any developing science. The interplay between the descriptive and the normative sides can be conceived, to be sure, as a sort of reflective equilibrium (Goodman 1965). But if one is dealing with a field where norms still have not had the time to get established, one runs the risk of extracting normative notions from any research trend that seems promising or fashionable at the time of the analysis. In this respect, von Eckardt analysis of connectionist systems is too oversimplified to fit her computational assumption. She accepts that symbol processing is a necessary condition for being a computer, which is an interesting and debatable normative feature. Yet she endorses, without much qualification, the view that artificial neural networks are

symbol-manipulating devices. She obviates, thus, the long and often sophisticated debate between classical and connectionist partisans.<sup>11</sup>

The computational assumption is rephrased, thus, as “the human mind/brain is either a general-purpose, stored-program, conventional computer or a special purpose, wired program, conventional computer or a connectionist computer, or some combination of these” (p. 140). Of course, an additional way of computing could be invented tomorrow, and a new “or...” should be added to the list already offered; or it may happen that the general-purpose view of a computer is deemed wrong for the ends of cognitive science, and erased from the catalogue of possible computational models of mind. But could we still consider the assumption as “computational”? How much revision can be tolerated before the cognitive science community accepted that “after all, the mind was not a computer”? The way von Eckardt frames the question, as something relative to available devices at the time, eschews many important problems that have to be faced in a principled manner. One of them is where to locate connectionist computational powers with respect to classical machines. Connectionist systems presented themselves as a computational link between different levels of description of mental processes. Most of the debate between connectionism and classicism was devoted to elucidating this claim. Adherents of connectionism argued that their systems offer a real possibility to bridge neuroscientific, computational and intentional descriptions. The point was not, then, whether the privileged unifying level of description was computational or not. The point concerned rather what kind of computational system ought to fill that level so that there could be effective links between levels. Connectionism embodied the promise of being,

<sup>11</sup> A typical instance is the exchange between Fodor *et al.* and Smolensky (cf. Fodor and Pylyshyn 1988; Fodor and McLaughlin 1990; Smolensky 1988, 1991). Von Eckardt mentions the literature but does not pay attention to it. There are other relevant classical-connectionist controversies, but they are too numerous to be mentioned here.

at the same time, representational and neurologically plausible. For the purpose of clarifying relations between cognitive theories we thus need to reverse the order of priority to formulate the computational assumption. The crux is not to judge if a certain device is computational or not, it is rather to determine whether a certain (putatively computational) system stands in suitable theoretical connections with other levels (particularly with intentional and biological descriptions). And the relation between classical and connectionist systems themselves is one of the relevant points in this respect. Offering an eclectic view of computation, von Eckardt simply does not allow us to see the relations between the computational species.

The problem is more apparent when we consider the role that she assigns to neuroscience. To be sure, this is an area where constraints on psychological hypotheses can arise. Yet, on her view the constraining character of neuroscience is relegated to a methodological assumption. This claim does not square with her insistence on attending to the actual practice in cognitive research, because many cognitive scientists would place neuroscientific constraints among the substantive assumptions. She distinguishes two major opinions in this respect: the “isolationist” stance, which vows for the autonomy of information-processing approaches from neuroscience, and the “constraint” stance, which defends a more restricting role of neuroscientific findings towards theories framed at a higher level. She tries to elaborate a position intermediate between both extremes. On the one hand, von Eckardt rejects the argument of multiple realizability, in the sense that it is not possible to conclude from it the non-existence of a constraining relation from neuroscience to psychology. The fact that at the moment there is no evidence of detailed psychoneural correlations does not tell us any fundamental fact about the nature of both levels. It is simply a reflection of the current state of technology, but nothing precludes the possibility that such correlations will appear in the future. On the other hand, she rejects a “bottom-up” approach that would take neuroscientific data as the starting point to develop a

theory. She opts, thus, for coevolution,<sup>12</sup> a strategy in which information-processing answers to the basic questions of cognitive science are developed on the basis of empirical findings from both the non-neural cognitive sciences and the cognitive sciences.

The defense of a coevolutionary approach does not tell us how the diverse empirical findings are to be related. Is what we obtain a unified theory? Or do we end up with different theories about the same subject matter? If, as it may happen, neuroscientific findings gain increasing weight in cognitive explanations, will that mean a reduction of the cognitive to the neural? If the answer is negative, how should we characterize the relation between them? Expressing it as one of coevolution is not to say much. We can also say, for example, that psychology will coevolve with mathematics, e.g., the development of further statistical tools will influence the way in which psychological facts can be researched and expressed, and conversely, the needs of psychological research can move the development of mathematics in a particular direction. Still, it would be preposterous to assert that there is any interesting intertheoretical relation here that we need to characterize. Thus we need to specify the notion of coevolution in a more precise way than “mutual influence”. We need to provide an account of how theories in different levels can (if they can at all) be integrated.

A relation of reduction is discarded if we assume von Eckardt’s “information-processing ineliminability thesis”. Even if we appeal to neuroscience, we cannot get rid of information-processing accounts. Thus, she allows the possibility that neuroscience could provide an information-processing explanation, as far as it were able to provide its own computational and representational hypotheses. Yet, it is not clear how this can be possible in a way other than the bottom-up approach that she already rejected. And it is not clear if the explanation in this case would be neuroscientific *and* information-level at the same time, or we would be “switching” back and forth between

<sup>12</sup> For one of the first formulations of the notion of coevolution see McCauley (1986, 1995). Von Eckardt’s view, however, is independent of it.

different descriptions. One possibility is that some discipline could afford a vocabulary that allowed facts from each level to “make contact”. Connectionism, as we already mentioned, is a view that promises the integration of neural constraints. Von Eckardt mentions the typical example is the 100-step constraint, which would restrict the neurological plausibility of computational models to those that perform a cognitive task in less than 100 computational steps. However, the demands that this constraint pose are quite unclear, because we lack any sound analysis of how to count computational steps, and how to map them onto neurological steps.

To sum up, von Eckardt’s reconstructive approach needs to be considerably reworked to ascertain the nature of intertheory relations in cognitive science. She sets forth an eclectic view which supports the desideratum that cognitive science will be able to profit from theoretical advances attained in any of the disciplines. But, in our view, she fails to motivate the view that computational descriptions are the appropriate common ground. It might be objected that we have mislabeled von Eckardt’s view by calling it “classical”. How can it be classical when she allows connectionist networks inside her picture and when she does not support an outright autonomy of psychology from neuroscience? This eclecticism should not lead to mistake. Von Eckardt admits connectionist machines as systems that may fill the computational part of the story, only inasmuch as they can be interpreted in symbolic terms, and she admits neuroscientific findings only inasmuch as they can be rendered in terms of information-processing constraints. Her picture is thus conservative. In the following section we examine a picture that, without endorsing a total break with traditional views, demands a thorough revision of them.

#### *4. The Connectionist Revision*

Von Eckardt’s reconstruction of the assumptions and aims of cognitive science hits on several of the aspects deemed crucial from a traditional computational perspective. Doing a similar reconstruction within a non-classical framework offers, in con-

trast, many more problems. On the one hand, it is reasonable to select connectionism, the family of non-classical computational systems that emerge in the 1980s, as the main alternative to von Neumann-style architectures.<sup>13</sup> On the other, interpretations of connectionist systems are much more varied than classical ones.<sup>14</sup> Hence, the lack of standardized assumptions is more pressing in the connectionist case. Thus we shall concentrate on what we regard as one of the most coherent construals of connectionism as a non-classical framework, developed by Horgan and Tienson (1989, 1992, 1993, 1994, 1996, 1997). It must be said that their aim is not simply to provide a coherent view of connectionism. Rather, they intend to articulate an alternative to classical cognitivism. Connectionism would be one of the candidates to embody their position, but they do not preclude that other systems may be better qualified for this role. In this section we will focus on those aspects of their proposal that pertain just to the relation between levels, as this is more relevant for the question of intertheory relations.

An important first step is that Horgan and Tienson see Marr's three-level distinction as a species of a more generic schema. Cognitive functions as defined in Marr's level 1 would be a subset of all possible cognitive state transitions. Algorithms at level 2 would be a subset of all possible mathematical state transitions, in particular, of computable transitions. But there are still a large number of non-computable transitions. Level 3, the implementation level, would remain basically the same in their schema. Just as it happened in Marr's distinction, each level in this generic schema presumes a functional decomposition at the immediately upper level, and it is multiply realizable in the immediately lower level. The idea is that each functional property of level 1 corresponds to a mathematical state at

<sup>13</sup> For a history of the connectionist boom, as well as different highlights of its psychological and philosophical import, see Bechtel and Abrahamsen 1991, Clark 1989, 1993, Horgan and Tienson 1991, or Ramsey *et al.* 1991.

<sup>14</sup> Connectionism has been recruited for their different purposes by holists, associationists, antirepresentationalists, eliminativists, direct perception theorists, orthodox computationalists, neurocomputationalists, and anticomputationalists.

level 2, where functional relations are mirrored by mathematical relations. Nonetheless, we have to take into account that there is more than one possible way to put mathematical flesh on functional transitions.

Horgan and Tienson assume the cognitive science tenet that mental properties are not the neurobiological (i.e., implementational) properties, but the abstract functional-organizational properties. In virtue of these properties, physical state transitions are systematically appropriate to the contents of the mental states they realize. Functional properties have a mathematical nature, and the classical view takes algorithms as the best way to capture them, because algorithms are specially apt to describe state transitions between formal, discrete, symbol-like representational structures. In this sense, an algorithm would be constituted by a set of formal rules that operate on representations. Yet, algorithm theory is only a part of mathematics. Other mathematical descriptions may be able to account for that intermediate level between mental and physical properties. In this way we may construe an alternative to classical computationalism, an alternative they label as *representations without rules*, in which: (a) each level is a different kind of explanation, (b) states in each level are realized on states of the immediately lower level, and (c) level-2 explanations are not necessarily algorithmic but they admit of a different sort of mathematics, such as dynamic systems theory.

The classical view, Horgan and Tienson contend, entails that cognitive functions have to be specifiable by means of general laws about cognitive states. These laws are realized in particular cognitive transitions, which can be specified by rules on the algorithmic level, so a computation can be determined. However, if there are no such general laws, then the cognitive function will not be tractably computable. In fact, they argue, this is the actual situation in cognitive functions. Generally speaking, psychological laws are “soft”. They express the behavior of a system in generalizations containing ineliminable *ceteris paribus* clauses. There is always the possibility that psychological-level factors exist preventing the *ceteris* from being *paribus*. These

laws are quite different from “quasi-exceptionless” laws typical in other sciences. In quasi-exceptionless laws it is admissible that some phenomena pose an exception to the law, but they are always phenomena that arise *outside* the domain of the law. By contrast, psychological exceptions come intrinsically from *within* the psychological domain itself. This is due to two important features of mental properties: the open-endedness of the representational capacity (in the sense that there are no prearranged limits to the uses it can be put) and the potential relevance of any thought or perception to any other (so that belief fixation depends on relations among large numbers of cognitive states).<sup>15</sup> We cannot have rules to capture cognitive transitions because it is impossible to specify in a tractable way every factor that can pose an exception, even when we constrain the cognitive domain we are going to study.

Soft laws should not mean a problem for psychology, Horgan and Tienson claim, given that they appear to be genuinely explanatory. If we tried to replace them with complex laws in which all the exceptions were listed we would miss important generalizations about mental life. These generalizations can be understood as “tendencies” of the system towards a particular state or solution. The tendencies reflect the strength of underlying defeasible forces that pull in different directions and that determine the behavior of the system in a non-additive way. So it is useless, in fact it is a mistake, to try to get rid of soft laws. They are as scientific as psychological laws need be. Still, they pose a problem for the classical view, since they block the possibility of a computational treatment of the generalizations expressed by them. Thus they need a different mathematical approach, one that could be afforded by dynamic systems theory. Connectionist systems could be thus characterized in this mathematical framework: their transitions do not generally conform to algorithmic relations but can be captured by dynamic

<sup>15</sup> This is a form of the multifaceted frame problem (cf. Pylyshyn 1987), very similar to the formulation that Fodor makes in relation to the modularity of mind (Fodor 1983), and a reason for his gloomy prospects for the study of the central systems.

mathematics. It is important to note that dynamic characterizations are more abstract than connectionist ones. The same dynamical characterization may be subserved by a number of neural networks, which allows us to explain how two individuals with different and particular brains can still stand in the same cognitive state.

The centrality of dynamical mathematics is a point that Horgan and Tienson share with other recent approaches, like Hardcastle's (1997). She also embraces the idea that mental functions are better modeled by a connectionist network. As we will see in the following section, computational neuroscience is one kind of bridge science that can help to relate brain functions—expressed in physiological terms—with patterns of mental properties—expressed in computational terms. She does not accept, however, the notion of levels and laws entailed by the “representations without rules” paradigm. There is not any explanatory distinction, Hardcastle contends, between the three levels proposed by Horgan and Tienson. When we outline the abstract, general functional organization of our nervous system, we are simply using one particular description of the system. A mathematical description at level 2 is just *another* description of the implementational level. Level 1 descriptions try to depict trajectories through *contentful* states, and different mathematical stories can be told about those states at level 2. However, these trajectories are just further descriptions, increasingly abstract and general, of the dynamics at level 2. Hence the distinction between levels collapses; every level is simply a different manner of characterizing the same interaction, a different way of accounting for the same set of trajectories that describes a certain sequence of neural activity. These levels of description do not correspond to different structures, whose theories might have a relation of reduction from the highest to lowest ones.

Yet, this interpretation of levels as levels of description can be assumed by the representation without rules scheme. Tienson (1997) agrees that we have three different groups of predicates, which are directed to different sets of systematic relations be-

tween states. Each level affords a predicate that can be applied to a certain state of the system. Thus, a neural network state can be described as a *representation*, as a point in an *activation landscape*, or as a *physical* state that encompasses a number of connected units. In addition, Tienison shares Hardcastle's agnosticism with respect to the ontological scope of the levels. Yet, a crucial difference between both approaches persists. In Horgan and Tienison's account there is a level whose description is properly cognitive, in the sense that it refers to intentional states. Psychological laws are formulated at this level, and the potential evolutions that drive a particular system from one state to another are better captured at a mathematical level of abstraction. Hardcastle, in contrast, does not have such a commitment with a "privileged" level. Her idea of a genuinely interdisciplinary theory involves the possibility of referring to the content of the states, even if we are simply describing the neurological substrate. A description that talked "purely" about contentful states (that is, leaving aside the physiological basis of those states) would just be an extremely abstract model of the very same domain, not a "qualitatively" different model.

There is a second difference that is even more difficult to save. Hardcastle assumes that psychology is analyzable in the same terms as any other science. So she rejects that psychological laws are special in any interesting respect. Neither laws of psychology nor laws of physics are exceptionless; they do not apply to all possible situations in the real world in which they hold, and it is not their task to do so. Scientific theorizing occurs at some distance from the real world, in an idealized sphere with well-defined parameters. All laws, not only psychological laws, are *ceteris paribus*, given that they abstract from details, simplify and quantify. The way in which psychology abstracts is certainly different from that of other sciences, because it is centered on a different set of regularities, and it possibly has to prune many more details than sciences like physics do. But the consequences of this labor arise in the *scope* of the laws so formulated —i.e., in the relation between the real world and the abstract system we have defined— *not* in the nature of

laws themselves or in the primary process by which they are obtained: abstraction and simplification.

We could extend this kind of criticism in the other direction. The point, to simplify it, is that cognitive affairs are so complex that they are not capturable in an algorithmic framework, with rules that mimic the transitions observable at the cognitive level. Yet, the same complexity appears when we look one level down, in the direction of biological processes. Neurophysiology, just as psychology, is not tractable computationally: there are simply too many interactions occurring at the same time to be able to capture them without falling into combinatorial explosion. Could we say that this meant a problem for the classical computational account? No, because the point of computation was to abstract over all that detail to obtain an intelligible generalization of whatever goes on at the physical implementation. Now, the interesting thing about this computational generalization is that it is also aimed at capturing the relevant parts and relations of the level immediately above it. As far as there are appropriate links with the lower and the upper descriptions, the computational story embodies the promise of providing a unifying account that connects all the levels in which our mental story can be told. If we are too strict with our computational description demanding from it either that it follows more closely the details of the cognitive level or that it stays closer to the biological account, we run the risk of losing precisely what we want from that description: its capacity to provide an intermediate bridge between theoretical accounts that are too far away to be related in a direct way. It might be that our mental life can better be modeled by means of dynamic systems mathematics, as Horgan and Tienson contend, instead of algorithmic operations. But this is an empirical question, whose answer is independent from the issue of whether psychological laws are soft. Even if psychological laws are as Horgan and Tienson portray them, algorithms could still provide the best mathematical model and their interactions could take charge of the complexity that is apparent in mental life.

Yet, it is possible to reject altogether the view that there is any “connecting level”. There may be, to be sure, links that relate hypotheses formulated in different theoretical descriptive frameworks. But we should not think that any of these links is *where* the unified explanation obtains. The global picture can be much messier, with explanations that employ cognitive, computational and biological elements related in different structural and functional ways. This is the view that we are going to examine in the next section.

### 5. *The Pragmatist Approach*

Hardcastle (1996) offers a pragmatist approach to cognitive science that rejects the idea of a privileged level of theorizing. This is the way in which we have to understand her initial statement that “cognitive science isn’t cognitive psychology”. One bias that she explicitly endorses is to consider brain science as an essential component in cognitive science. Yet, this does not entail that biology ought to be privileged in theory building, since she qualifies her position with a second explicit bias: to regard cognitive scientific theories as strongly interdisciplinary. Her biological bias is offered as a counterpoint to classical views, like von Eckardt’s, which ignore neuroscience and gratuitously constrain the topics. But this primacy of the brain is not understood as entailing a reductionism of the rest of the levels to biological relations and explanations.

Hardcastle accepts that cognitive science is, in general, the study of information processing. Even if this definition seems to differ very little from the classical view that she criticizes, a look at the heterogeneous list of disciplines that she includes in cognitive studies evidences the comprehensive interpretation that she gives to the notion of information processing. The list includes anthropology, biology, computer science, engineering, linguistics, mathematics, philosophy, psychiatry, psychology, neuroscience, and sociology. We find this eclecticism very questionable. To say that mathematics, sociology or engineering “share an interest in explaining how we process information”, and that they differ merely in the questions they pose and the

framework they set, reflects an odd conception of those disciplines. It seems that some of them provide a technical apparatus that can be used in studying information processing, rather than study processes themselves, while others study phenomena that, no doubt, arise from cognitive interactions but that receive influences which are at several removes from cognition. We should not say that, if cognition is the core of human mental life, and as human affairs are practically involved in everything, then we ought to see cognitive science spreading everywhere. There must be constraints about what can and cannot be included in it, even if the boundaries are certainly diffused.

Hardcastle's project, as other modern epistemological approaches (especially those with a naturalistic slant), wants to be at the same time normative and descriptive. The normative aspect lies in the idea that interdisciplinarity is regarded as a necessary feature, not a mere desideratum, of cognitive theorizing. This implies that a cognitive theory that does not embrace elements from different levels will leave out significant parts for a complete explanation, that is, it will not be a good (or an optimal) theory. The descriptive aspect, by contrast, comes from the analysis of paradigmatic cognitive theories. Descriptions will reinforce the normative project as long as we are able to show that the best cognitive paradigms are in fact strongly interdisciplinary.

When it comes to choosing theoretical paradigms there is an obvious problem of circularity that must be faced: that the examples have been chosen precisely because they exhibit the feature that we want to turn into norm. A second problem is that, even if we can guarantee the independence of the samples (e.g., calling on a consensus among researchers about which cases count as the best paradigms of cognitive studies),<sup>16</sup> it remains possible that the alleged normative feature is not but accidentally shared by the paradigmatic cases. If two theories  $T$  and  $T'$  are both good and both share property  $P$ , it does not

<sup>16</sup> Part of the success of Marr's three-level proposal comes from the relative consensus that his work on vision enjoyed for a time as a paradigm of cognitive research.

follow that it is *P* that makes them good theories. Thus, if one were to take this line to argue for the normativity of interdisciplinarity, then one would have to show that interdisciplinary theories are superior *because they are interdisciplinary*.

As for ontological questions, Hardcastle contents herself with the assumption that mental properties are simply physical properties, without taking a stand on the long-debated issue regarding mind-body identity theories. She simply contends that even if mental properties may be very complex or abstract, they exist at the same level of organization as some physical properties. The lower level structures are primary because they appear to be ontologically prior, but high level properties cannot be defined in those low-level interactions. This is what gives her project its non-reductionist character: the lower level is not explanatory prior or metaphysically superior to high level mental phenomena. It is *epistemic power*, i.e., the capacity to answer our questions, that determines whether the entities we posit in our theories must be maintained or rejected. Such questions, on their part, are driven by pragmatic factors, so the kind of explanation we adopt will depend, among other things, on the audience and the background theories we hold. This move from an ontological stance to an epistemological one is reflected in her verdict that “the question of what is *really* out there is no longer apropos” (Hardcastle 1996, p. 30).

Although the idea of a privileged level does not square with her eclecticism, Hardcastle claims that there is a level which is fundamental for explaining cognition: the middle level of organization in the brain. Here, for instance, is where we may observe patterns of activation that would constitute an appropriate middle level to study cognitive properties. Yet, we must not think that this favors biological approaches to cognition: the middle level is fundamental inasmuch as it allows in our explanations a convenient trade-off between amount of detail and useful generalization. In Hardcastle’s view, level divisions have only an epistemological significance, and cognitive theories still involve a range of questions that embraces several disciplines.

Accordingly, the different ways of formulating levels depend just on *our* different approaches to the phenomena we want to study, and this, in turn, depends on pragmatic considerations. However, many cognitive scientists, especially those of a connectionist inclination, portray levels in an ontological way, especially when it comes to describing structures that get associated to compose another entity. On the other hand, there are those that think that there is a continuity between cognitive and neural processes. How to characterize something as an algorithm or as an implementation will depend on the description we adopt, so that implementation details must figure within cognitive explanations. In this respect, Patricia Churchland (1986) takes it as controversial that we can describe the complexity of brain organization in terms of Marr's three levels. How many levels there are and what they are cannot be determined beforehand. She regards as possible levels, for instance, cell membrane, cell, synapses, cell assembly, neural circuit, or neuron. Each level can be seen both functionally and anatomically, so this last distinction is also relative. In addition, the symbolic and subsymbolic levels posited by Smolensky (1988) can be read structurally to account for the relation between classical and connectionist models: each of them would be dealing with a different level of explanation, which would accordingly correspond to a different level of reality.<sup>17</sup>

The paradigm chosen by Hardcastle to illustrate her views on cognitive theorizing is the dual memory hypothesis. The central idea is that there are two information retrieval systems that exhibit different properties when considered from diverse perspectives, such as cognitive psychology, developmental psychology, neurophysiology or clinical neurology. They examine the same general phenomenon, but pose questions that do not fit together easily. What keeps these paradigms together is the existence of some methodological similarities. In particular, all of them use different versions of the same basic task, which provides an

<sup>17</sup> Indeed, Smolensky's proposal is similar to a micro-macro relation between classical and connectionist levels, so that the former could be reduced to the latter.

intermediate description that captures what is common across several observations. It seems now that the idea of a privileged level reappears in its abstract formulation, i.e., as the common ground. Yet, Hardcastle warns against any temptation of obtaining a general framework that comprehends all the distinct models. Cognitive theories will always be *messy*, in the sense of mixing structural and functional elements from more than one level. It is possible for a discipline to yield an *explanatory extension* (Kitcher 1981) for another one, but the lack of structural isomorphisms between them will preclude the possibility of obtaining a single basic formalism. Hypotheses are linked by relations of similarity. Family resemblance in structure allows the joining of a number of models in a single “smeared out” interdisciplinary theory, but we should not expect that they “collapse into a single overarching account” (Hardcastle 1996, p. 139).

This scheme honors interdisciplinary pluralism. However, can it also warrant the unification of the different theories? It is difficult to resort to the notion of similarity for this goal. Unless we make this notion more precise it will always be at the mercy of pragmatic factors that establish what respects are relevant for the similarity relation. And it is obvious that transitivity cannot be guaranteed at this point. A model  $M$  may be similar to a model  $M'$  (according to a parameter  $P_1$ ) and  $M'$  may be similar to  $M''$  (according to a parameter  $P_2$ ) and yet  $M$  and  $M''$  may have no resemblance at all (because  $P_1$  and  $P_2$  involve different factors). Thus, Hardcastle needs another kind of force to keep theories united. She finds this glue in *bridge sciences*.

A bridge science allows us to establish connections between different disciplines with diverse vocabularies. There can be different bridges to link different pairs of disciplines. An example of bridge science between psychological and neurological data is *evoked response potentials* (ERP). ERPs can be tied, on the one hand, to (gross) anatomy and physiology and, on the other, it can assume the descriptive psychological categories that individuate in terms of “perceived experiences” and the like. Yet,

this example is still unclear with respect to the actual role of a bridge science. Just as bridge laws in classical reductionism (Nagel 1961) were not laws in the same sense as theoretical laws, a bridge science does not seem to be strictly a science, but a sort of methodological auxiliary. If the bridge law did not belong to a particular science and it specified the deductive relations between the reducing and reduced theories, a bridge science would specify the (non-reductive) relation between descriptions of different levels. In classical reductionism the relation was given in logical terms, but Hardcastle cannot avail herself of this idea. So her most likely alternative is to resort to similarity relations again. This is more evident if we take into account that Hardcastle wants to stress the idea of science in 'bridge science'. ERP descriptions, she argues, count as functional decompositions of higher level psychological theories and provide a theoretical framework distinguishable from the "bridged" sciences. But then the idea of bridge science does not add anything to the idea of similarity because we need to come back to this concept to account for the relation between the bridge science and each of the sciences it connects.

In short, Hardcastle approach is afflicted by the weaknesses that pragmatist approaches generally share: the lack of appropriate constraints and, hence, the risk of the "anything goes". Interttheory relations have to be loose enough to allow a plurality of pragmatic factors to influence them. But the outcome is unrestrained eclecticism. The insistence on messing up functional and structural elements is not a negative point *per se*. It may be true that middle level descriptions are "neither purely functional nor structural" (1996, p. 103). But if it is here where most explanatory work occurs, it seems that we need some constraints in order to determine what counts as an appropriate middle level, and that we need a stronger account of the relationship between this level (which is, despite Hardcastle's claim, a privileged level) and the levels which lie above and below it. The forces that keep a theory unified are not strong enough, so it is unclear that we can talk about a single theory. In the following section we will examine a proposal that follows the

opposite direction, i.e., it imposes a restricted corset that sends cognitive science to the reductionistic path.

### 6. *The Reductionist Account*

If reduction has often lured philosophers of science with its promise of the unity of scientific knowledge, its charms are almost impossible to resist in an area in desperate need of some cohesion between its disparate associate disciplines. Alas, if reduction presented problems for hard-core science, they increase when it is transported to the realm of cognitive science. Figure 1 shows how classical reduction would depict “ideally” the relations between three putative cognitive levels.

The elements individuated at the upper folk psychological level (beliefs and desires) have correlates in the elements distinguished in lower levels (states in an information processing system and neuronal states). Correspondingly, the arrows that link the elements within each level represent theoretical relations that also have their correlates in the relations that exist between elements of the other levels. There are rules of correspondence that provide the links between these levels, so that we can establish a deductive relation between the theories at the different levels.

We call this picture “ideal” because it corresponds to what intuitively philosophers have demanded from reduction. Nevertheless, viewed from the perspective of classical intertheoretical reduction it is difficult to see how this picture would work. To begin with, there are some general problems with the classical account. The best known problem lies in the fact that since the reduced theory is deduced from the reducing one, if the former is false (as it will almost always be the case) then the latter will also be so by *modus tollens*. To this problem, common to any reductive scheme, we can add others that are more specific to cognitive reduction. One is multiple realizability, which in our diagram is illustrated by the fact that the arrows connecting different levels will be in a one-to-many relation. Another has to do with the absence of laws, both intra and intertheoretical: (i) the status of folk psychology as a “proper” theory is under

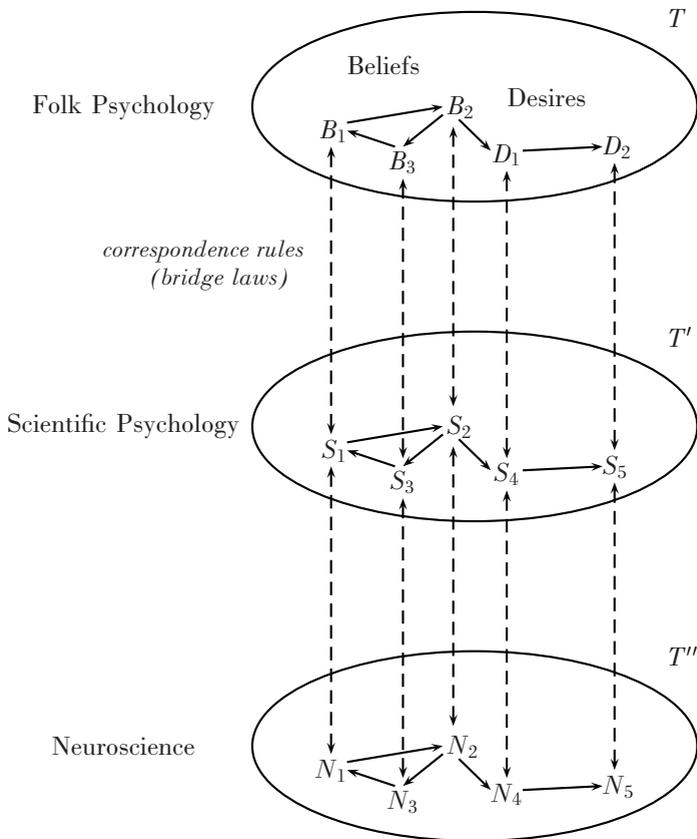


FIGURE 1. Interttheory relations: "Ideal" classical schema.

discussion;<sup>18</sup> (ii) even if theories of scientific psychology seem better situated for reduction, psychophysical laws linking psychological and neurological predicates have a dubious status as well. All this challenges the possibility of establishing any sort of bridge laws. Thus, if you don't have two theories (folk and scientific psychology), or you don't have laws connecting them, you cannot even try to consider the possibility of reduction.

In a classical paper, Schaffner (1967) offered a classification in *styles* of intertheory reduction that he called "paradigms":

1. The Nagel-Woodger-Quine (NWQ) paradigm: it is a model of *direct reduction*. The basic terms (and entities) of a theory are *equated* with the basic terms (and entities) of another theory, and the axioms and laws of the reduced theory are derivable from the respective ones of the reducing theory. As it may be the case that some terms in the reduced theory do not occur in the reducing theory, it is necessary to add other statements to the latter, so that those terms in the reduced theory become associated with combinations of terms that belong to the vocabulary of the reducing theory. (An example of this is the classical reduction of thermodynamics that Nagel commonly uses.)
2. The Kemeny-Oppenheim paradigm: it is a model of *indirect reduction*. Rather than deriving a theory *T* from a theory *T\**, it requires the ability to obtain identical observable predictions from both theories. It also requires that a greater number of phenomena can be predicted from the reducing theory as a condition to warrant an asymmetrical relation between both theories. (An example of this reduction seems to be Lavoisier's theory of oxidation, which predicts all the phenomena that were observable by the phlogiston theory.)

<sup>18</sup> There has been a large controversy on this issue. Paul and Patricia Churchland (1989; 1986) figure among the hardest defenders of folk psychology as a theory, although they consider it a *bad* theory, demanding thus for elimination. This view has received lots of criticisms rejecting such a theoretical status. For further discussion see also Davies and Stone 1995.

3. The Popper-Kuhn-Feyerabend paradigm: this model is aimed at the clarification of the dynamic relations in science, that is, the relations between predecessor and successor theories. This proposal can be interpreted as an argument *against the possibility of reduction*, rather than as a model of reduction. Yet, Schaffner thinks that it is possible to make a positive version of the argument. It is not required that  $T$  be derivable from  $T^*$ , but  $T^*$  must be able to explain the success of  $T$ , *why*  $T$  works. Some derivability is even possible if we add additional premises and demand just approximate experimental results (for instance, the Galileo-Newton cases).
4. The Suppes paradigm:<sup>19</sup> we can obtain a reduction *iff* we are able to prove that, for every model  $M_2$  of the reduced theory  $T$ , it is possible to find a model  $M_1$  of the reducing theory  $T^*$ , so that it is possible to construct a model  $M_1^*$  (which can be identical to  $M_1$ ) that is isomorphic to  $M_2$ . The problem here, as Schaffner points out, is that Suppes offers no general definition of isomorphism, on the grounds that it is a difficult task.<sup>20</sup>

Because of the difficulties of classical account, the few and very recent attempts at tackling the problem of reduction in cognitive science from a general philosophy of science perspective adopt precisely the semantic view characterized in (4). Hardcastle (1996) leaves undecided whether her semantic view approach is essential to her arguments, as well as which of the

<sup>19</sup> See Suppes 1957.

<sup>20</sup> Schaffner emphasizes that this is not a trivial issue. The problem lies in the fact that if we take whatever of the typical definitions of isomorphism, it is easy to show that Suppes' account amounts to a weak version of the NWQ's paradigm. In fact, it is so weak that it is difficult to regard it as a reduction. Schaffner concludes saying that the Suppes paradigm is so weak that it allows different theories to have the same formal structure (e.g., thermodynamics and hydrodynamics), and yet it makes no sense to attempt a reduction between them. In this account we could say that isomorphism is a *necessary* but not *sufficient* condition for reduction.

different versions she takes to be the right one.<sup>21</sup> Bickle (1998), in contrast, takes sides with the structuralist view of reduction. The reason seems to be that while Hardcastle endorses the generalized (in the semantic view) tendency to accept the primacy of epistemology over ontology,<sup>22</sup> Bickle is convinced that there *is* a real mind-body problem. His initial assumption is that contemporary antireductionist physicalism is an unstable program that fluctuates between reductionism and property dualism. *Prima facie*, the antireductionist physicalist cannot have it both ways. Classical reductionism sees the reduction relation as a *deduction* of the reduced theory ( $T_R$ ) from the reducing theory ( $T_B$ ). Given that  $T_R$  and  $T_B$  are expressed in different vocabularies, we *need connecting principles* between terms in  $T_R$  and terms in  $T_B$ . These principles must express, at least, *accidental coextensions* (many regard them as nomic) between the related terms.<sup>23</sup>

<sup>21</sup> If we look at the classification of semantic approaches in terms of the mathematical tools they use, set theory or state-space theory, Hardcastle exploits the former. Yet her examples do not make clear whether she is using Suppes' version or the structuralist one: "Let me emphasize, though, that I adopt the semantic view in this text not because I think it is inherently superior to the more traditional models (though I do) —indeed, every thing that I claim can be translated into any view of theory you might have with greater or lesser ease— but, instead, because it highlights certain features of theories that are important for the discussions above" (Hardcastle 1996, p. 175). We ultimately suspect that Hardcastle is *not* interested in the reduction problem. She takes ontological monism for granted, but she regards it as a waste of time to clarify this. Instead, she is more interested in using the semantic view to characterize cognitive science theorizing, where interdisciplinary theories are developed. But contrary to what Hardcastle assumes, the massive practice of *messy* theorizing in cognitive science does not invalidate reductive issues; it simply adds an interesting turn to them.

<sup>22</sup> The notion of 'epistemology' employed in some structuralist quarters is quite controversial. It pays little attention to normative problems of epistemic justification and, in general, it abandons the ontological questions. There are exceptions (such as Suppes' and Giere's), but they are not generally developed in the structuralist framework but in other versions of the semantic view.

<sup>23</sup> Some authors affirm that such coextensions should be definitional, but others defend the view that accidental coextensions are good enough. For example, Boyd (1991; 1999) and Millikan (1999) claim that coextensions should be accidental, at least in the case of historical kinds. This accidental character

The problem is stated as follows. Let us agree that two typical antireductive principles hold: the anomalousness of both mental and multiple realizability. Then it follows that classical reductionism cannot work, because both positions assume the impossibility of psychophysical laws. A way out might be some weaker relation between theories located at the physical and mental levels, respectively. It is in this direction that Bickle will go in his structuralist approach.

One of the main reasons that Bickle offers in favor of structuralism as a reduction schema to deal with the mind-body problem is his analysis of Hooker's criticism (1981) of classical reduction. Hooker deals with one of the already mentioned basic problems of classical reduction: the falsification of the reducing theory, by *modus tollens*, when the reduced theory happens to be false. The common solution of complementing the reductive complex with limiting conditions and assumptions will not suffice for the task. Schaffner-style solutions, according to Bickle, will not do either: instead of deriving the original reduced theory ( $T_R$ ), Schaffner proposes to derive a corrected version ( $T_R^*$ ) constructed from a reducing complex that uses the reducing theory along with a number of conditions. This maneuver allows us to escape the problem of the falsity of the reducing theory, but it is useless in cases in which reduction involves ontological elimination.

If the solutions above are not acceptable, then we seem to be forced to accept a Kuhn-Feyerabend reduction, i.e., there would be no real reductions but just a replacement of some theories by others. Bickle, however, does not think this is necessary if Hooker's proposal is adopted: instead of regarding reduction as a deduction of a structure specified in the reduced theory's vocabulary (even if it is modified *à la* Schaffner), we can assume that what is deduced is a structure specified in the reducing theory's vocabulary, along with certain restricting conditions.<sup>24</sup>

doesn't deprive them from being nomic and robust, in the sense that they provide a basis for induction.

<sup>24</sup> In fact, Schaffner's proposal, which attempts to accommodate the most relevant intuitions from the four reduction paradigms he reconstructs, is inter-

In this way we obtain a proper subset (a set of theorems) of  $T_B$  that, following an analogy relation, turns out to be *relevantly isomorphic* to  $T_R$ . We have to take into account that the pairs related by the analogy relation are not synonymous, or coextensive, or synthetic identities in the usual sense. However, Bickle does not provide any clue for characterizing such a relation. Moreover, it seems that he is not interested in it, because what he wants to defend is the view that in the structuralist approach to reduction “there is room for everyone”. If the restricting conditions are not wildly counterfactual and most laws in  $T_R$  have syntactical analogues in the corrected version  $T_R^*$ , then we can hold that there are identities between the referents of the terms in the intertheory pairs. In other words, if we have reductions that do not involve large corrections of  $T_R$ , then we will have ontological identities. In the opposite case we can have eliminations. There is in fact a wide spectrum that ranges from smooth reductions to bumpy ones, and that is mirrored in the ontological plane, drawing a continuous line that begins with identity and ends in elimination. Figure 2 illustrates this situation.

What Bickle is proposing, in effect, is a criterion of “ontological pruning” that justifies the elimination of properties inasmuch as the theory in which they appear stands in a bumpy reduction relation with another, more basic, theory. This criterion can be applied to the mind-body problem because this, he says, is primarily about how to relate two theories (folk psychological and neuroscientific) that deal with human behavior. Ontological consequences are dependent on this preliminary question. There is a long debate about the theoretical status of folk psychology, but Bickle does not provide any additional argument to sustain his position. So we may interpret his conclusions as conditional, subordinated to the claim that folk psychology is a theory in the required sense. He suggests that folk psychology is probably more than a theory, but no less than that. His exact position is, in general, difficult to ascertain: it is not clear whether the

preted by several authors exactly in the same sense in which Bickle interprets Hooker. See, e.g., Bechtel 1988 and Rosenberg 1994.

TYPE OF REDUCTION

PERFECT SMOOTHNESS		LIMIT BUMPINESS
<i>light</i>	<i>theory of gas</i>	<i>phlogiston</i>
PERFECT RETENTION	REVISION	COMPLETE ELIMINATION

ONTOLOGICAL DECISIONS

<i>IDENTITY THEORIES</i>	<i>SYNTACTIC ELIMINATIVISM</i>
<i>(Smart)</i>	<i>(Stich)</i>
	<i>NEUROSCIENTIFIC ELIMINATIVISM</i>
	<i>(Churchland)</i>

FIGURE 2. Type of reduction determines ontological decisions.  
Adapted from Bickle 1998.

candidate to be reduced by neurophysiology is folk psychology, or specific theories from scientific cognitive psychology. If we consider his initial motivation, and his insistence that the mind-body problem must be handled as a problem of intertheory relation, in which we must pay attention to real scientific practice, we may opt for the second interpretation.<sup>25</sup> Yet he often poses the problem in terms of folk psychology, as a theory in itself, to be reduced to neuroscientific theories couched in neurophysiological terms. A possible reason for this might be that with his “there is room for everyone” argument he wants to incorporate solutions, afforded by the philosophy of mind, for the problem of intentionality.<sup>26</sup>

<sup>25</sup> The cases Bickle uses to apply the structuralist frame of reduction are indeed of this sort. For example, the consolidation Long Term-Potentiation (LPT) Link. See Bickle 1998 and 2002.

<sup>26</sup> As an anonymous referee suggests, even if Bickle’s candidates for reduction were specific theories from cognitive psychology, it is unclear that this would leave out intentionality. We agree with this, but we would like

The advantage of reformulating the mind-body problem in the intertheory parlance is, Bickle says, that every traditional solution to the problem has its corresponding position in the catalogue of predictions about the “reductive fate” of folk psychology. So identity theories correspond to smooth reductions, while the various eliminativist versions correspond to reductions that are merely bumpy. This recipe looks suspicious. It seems that it forces psychological theory to face the following dilemma. Suppose that the reduction from psychology to neuroscience is bumpy. Then, according to Bickle’s schema, we have the right to eliminate psychological entities. Now suppose that the reduction is smooth. Then we can keep the entities, but they are explanatorily redundant. In other words, psychological notions are condemned either to the fate of phlogiston (they refer to nothing), or to the fate of light (they refer to something, but in terms that are theoretically impotent). Bickle’s view is that the relation between psychology and neuroscience will be located at some point between those extremes, implying some amount of revision of the upper theory. This way it might be possible to safeguard some theoretically interesting role for psychological description. Yet, it is certainly an odd position to affirm that the future of psychological theories depends on their capacity to occupy this middle position. Rather, we think that ontological decisions are independent of reduction decisions. If psychology were only bumpily reducible to lower levels it could still be the case that we want to keep its ontology, on the basis that it captures generalizations that are unavailable in those lower levels.

Even if one does not accept Bickle’s dilemma, his model of intertheoretical reduction may still be satisfactory. The key idea is that the analogy relation between the model obtained from  $T_B$  and the model of  $T_R$  that we want to reduce can have different degrees of strength. As far as the analogy is strong it will be possible to talk about a reductive relation

to note that while there is not much discussion that folk psychology appears customarily in intentional vocabulary, there is a long-lived debate on whether cognitive scientific theories can dispense with talk of contentful, intentional states (see, e.g., Stich 1983).

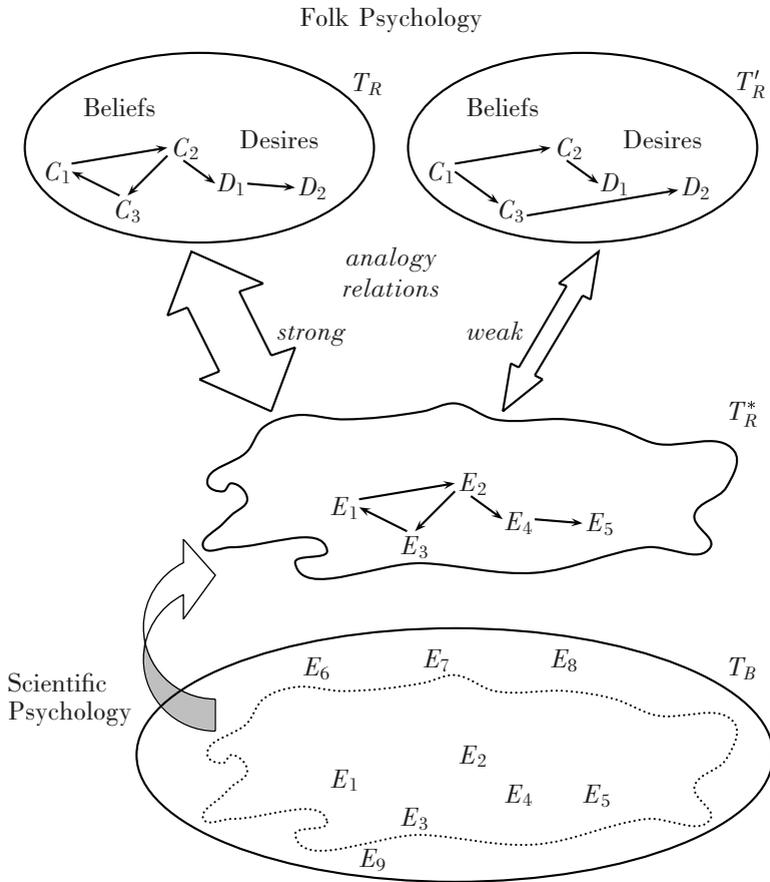


FIGURE 3. Analogy relations between a model of  $T_B$  and two possible models of  $T_R$ .

between the theories. In the limit, it will get closer to the identity relation. Conversely, a weak analogy does not sustain reductive links. Figure 3 offers an illustration of the relation between a model (that belongs to  $T_R^*$ ) of scientific psychology ( $T_B$ ) and two possible models of folk psychology ( $T_R$ ). Notice that the analogy is established between whole structures, and not between particular elements.

Structural relations are much more similar in the left part of the figure. So the folk psychological model on the left will be more smoothly reduced to the scientific psychology model. In fact, the idea is a bit more complicated, because relations are established between models of the structure  $T_R^*$  and those models of  $T_R$  that are confirmed empirical applications. Figure 4 is aimed at illustrating this extension. There is a relation  $\rho$ , which constitutes an ontological reductive link between the models. The link can be homogeneous or heterogeneous, or both. If the link is homogeneous then relations of identity are likely, while a heterogeneous link will support different degrees of model overlapping.

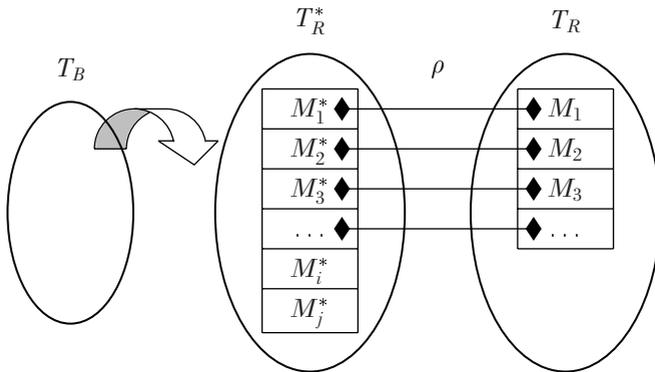


FIGURE 4. Relations between models of  $T_R$  and  $T_R^*$ .

Among the problems of the proposal, first we must note that the analogy relation, which is relatively weak and undefined, does not sustain the *transitivity* of reduction. Hence, it might

be the case that folk psychology (or portions of it) becomes reduced to scientific psychology theories, and the latter theories in turn to neurophysiology, while the reduction of folk psychology to neurophysiology turns out to be implausible. Bickle wants to legitimize his proposal by showing that there are parts of cognitive science that are reducible in his terms. His paradigm is associative learning. He describes the important parallelisms between learning processes described at a cellular level in the *Aplysia mollusk*, and learning as described in conditioning theories. Those parallelisms are evident in the formalisms of the laws of the respective levels. His idea is that it is admissible to suppose that higher learning patterns evolved from the lower ones, so that the same basic principles are shared. Leaving aside the plausibility of this assumption, we can raise two doubts about this paradigm. First, is it a clear and genuine case of reduction? Second, if it is, are its features exportable to other cognitive theories? With respect to the first question, Schouten and Looren de Jong (1999) argue that Bickle's examples are not reductions but, at most, explanatory extensions. Results from biological theorizing exert a selective pressure on psychological theorizing. But the latter also exerts a top down pressure on possible answers to biological questions. Relations between neuroscience and psychology are better portrayed in the coevolution model proposed by McCauley (1986, 1995). According to it, two disciplines in vertical relation do not compete, but constrain each other.

In our view, it is not clear that coevolution is incompatible with reduction. For Bickle, reduction is a relation that occurs *at the end*, when the theories are fully developed. In this case, the process of development, be it coevolution or any other, has relatively little importance. What matters is that at the end of the process it is possible to get an appropriate relation of analogy. In fact, Bickle endorses a "principle of mutual coevolutionary feedback" that fits McCauley's model. It may be a mistake to offer associative learning as an actual example of reduction, but it might still be a potentially reducible one after

further theoretical developments. However, our second doubt persists, are consequences in this field transportable to other cognitive areas? To this end we ought to show it is possible to link those areas with associative learning theories. Unfortunately for Bickle, there are no good prospects for this project. We would need to show that areas such as language or reasoning can be regimented in the categories established by conditioned learning theories. Yet, it was precisely the difficulties of such a project that conspired to make the behaviorist paradigm fall. As Richardson (1999) remarks, Bickle's model does not involve representations, only intermediate states. These notions are not equivalent, and no current alternative that accounts for high-level cognition in non-representational terms exists.

There is a further problem with Bickle's account. As we said, his solution is an attempt to fit Hooker's proposals in the structuralist model. The corrected version  $T_R^*$  has to be included into the set of potential models for  $T_B$  ( $T_R^* \subseteq Mp(T_B)$ ) and then everything follows the standard version of reduction in structuralism. But this solution is contentious. It is well known that structuralism finds the conditions of classical reduction (connectability and derivability) too rigid, and that they are replaced by something weaker. "Connectability" is replaced by a looser relation such as "the global correspondence between the models ( $Mp$ ) of  $T_R^*$  and  $T_B$ , and the respective mappings". "Derivability" is replaced by the notion that whenever there is a mapping that meets the laws of  $T_B$  (i.e., that is extendible to some  $M$  of  $T_B$ ), and that additionally meets some specific conditions (i.e., that is extendible to some  $M$  of an specialization of  $T_B$ ), then the correlated structure of the mapping in  $T_R$  will meet the laws of the reduced theory  $T_R$  (i.e., it will be extendible to an actual model  $M$  of  $T_R$ ).

One of the basic motivations of the structuralist view was to solve the issue that classical reduction demanded, as a previous condition, to have axiomatized theories. Yet structuralism needs at least to define the set-theoretic predicate of the candidate theories  $T_R$  and  $T_B$ . If this is so, it is difficult to see how to define the set-theoretic predicate of folk psychology. Assuming,

and this is too much to assume, that we obtain it, Bickle has to say how to outline the T-theoretic constituents from the potential models of such a theory. This is clearly preposterous. Thus we are inclined to think that what Bickle tries to do is only applicable, if at all, to the attempt to reduce cognitive scientific theories to others described at a neurophysiological level. He must forget about folk psychology. But in this case he must also forget about the problem of *intentionality*, that is linked to this psychology.

If the previous objection is right, then it also undermines Bickle's effort to show that Davidson's arguments do not affect his reformulation of classical reductionism, given that it does not require bridge laws. On the other hand, the arguments he offers in relation to the problem of levels and multiple realizability make more sense, but the fact is that he could maintain them outside the structuralist view. Bickle adopts Kim's analysis that explains that multiple realizability leads to specific or local reductions of the functional states to their particular realizers, just like temperature reduces to kinetic energy of molecules in the case of gases, but reduces to (is identified with) different things in the case of solids or plasma. However, none of this is new, and Bickle does not make clear how it applies to his position.

To sum up, Bickle's reductionism is *revisionist*. It is neither *absolutely conservative* nor *totally eliminativist*. Something is preserved and something is eliminated. But this statement is vacuous unless we make precise the proportion of conservation and elimination. Otherwise it simply means that psychology and neurophysiology will coevolve, will influence each other and will converge at some points. This is something nobody would call into question. Yet, in which sense can we say that "revisionism as a consequence of theory coevolution" is a sort of reduction? The end product of a process of coevolution of theories is a rather different thing: not the old theories, the candidate for reduction and the reducing one, but a new theory. Hence, Hardcastle's and Bechtel's reservations concerning reduction are understandable, and Bickle's final result is quite similar to their respective proposals. Finally, in the absence of stronger

philosophical arguments, it seems that Bickle's reduction could be compatible with some forms of dualism, inasmuch as it is helpless with respect to the problem of *intentionality* in folk psychology. This problem is common to all reductionist proposals that deal only with intertheory relations, and that do not want to discuss ontological issues, not even about the kind of realism that we have to assume for our theories.

### 7. *Conclusions*

Intertheoretical relations in cognitive science are difficult to elucidate. The levels involved are rather diverse, there are metaphysical problems to be solved, vertical and horizontal relations are often easy to confuse, and mental phenomena constitute a complex subject matter. The proposals we have examined face different problems. There is an aspect that is common to all of them (especially to the last three), i.e., their confidence in the possibilities of connectionism to provide a link between the representational and biological spheres. Leaving aside the problem of the causal efficacy of connectionist representations, that are usually construed as partitions of vector spaces, it must be noted that the relations of artificial neural networks with the levels immediately above and below them is seriously underdetermined. The fact that we have a system structurally similar to neurological systems and that also performs computational operations like a cognitive system does not entail that we *ipso facto* obtain a nexus between both levels. To this end we need an additional element: an explanation that tells how the levels are related. Connectionist models do not constitute by themselves, to use Hardcastle's term, a bridge science. The bridge is given by the models *plus* the principles that rule the projections from the elements of a model to the elements of a different level. The difficult task, the task in which none of the accounts manages to succeed, is to give an adequate account of these principles. The relative lack of success of connectionism to provide this bridge—a promise that explains much of the interest that they initially arose—is due to its inability to give principles in both directions: on the one hand, the deficiency of a psychoseman-

tics that defined the relation between networks and intentional properties; on the other, the lack of definition in the relations between structural properties of artificial and natural networks, fluctuating between macro and micro interpretations.

All the proposals here examined try to safeguard pluralism in cognitive science while establishing a framework of intertheory relations that goes across different levels. However, there is an unstable equilibrium between interdisciplinarity and theoretical unification. Maneuvers to save the former seem to lead to multilevel theories with little vertical cohesion. Solid unification tends to lead to a reductionist path. The most promising way, we suggest, is to be found in naturalistic approaches to representation. It is here, after all, where the link between psychology and biology is located. Constraints for interdisciplinary assembly must stem from it. Vertical theories of representation (the relation between vehicle and content) and horizontal theories of representations (the relation between vehicles themselves) need to be developed in such a way that they can constrain each other. Psychological, computational, and biological theories ought to be construed representationally, and the relationship between their respective representational systems would give an important clue to understanding the relation between cognitive theories.<sup>27</sup>

## REFERENCES

- Abrahamsen, A., 1987, "Bridging Boundaries versus Breaking Boundaries: Psycholinguistics in Perspective", *Synthese*, vol. 72, pp. 355–388.
- Bechtel, W., 1994, "Levels of Description and Explanation in Cognitive Science", *Minds and Machines*, vol. 4, pp. 1–25.

<sup>27</sup> This paper was written as part of research projects UPV/EHU 109.109–HB 078/99 and BFF2002–03842 of the Spanish Ministry of Science and Technology. Authors' names figure in alphabetical order. Fernando Martínez Manrique was granted a scholarship BF100.44 from the Education, Universities, and Research Department of the Basque Government. We would like to thank S. Silvers and an anonymous referee for their helpful comments.

- Bechtel, W., 1988, *Philosophy of Science. An Overview for Cognitive Science*, Lawrence Erlbaum, Hillsdale.
- Bechtel, W. and A. Abrahamsen, 1991, *Connectionism and the Mind*, Blackwell, Oxford.
- Bechtel, W. and R.C. Richardson, 1993, *Discovering Complexity*, Princeton University Press, Princeton.
- Bickle, J., 2002, "Concepts Structured Through Reduction: A Structuralist Resource Illuminates the Consolidation — Long-Term Potentiation (LTP) Link", *Synthese*, vol. 130, pp. 123–133.
- Bickle, J., 1998, *Psychoneural Reduction: The New Wave*, MIT Press, Cambridge, Mass.
- Boyd, R., 1999, "Kinds, Complexity, and Multiple Realization", *Philosophical Studies*, vol. 95, pp. 67–98.
- , 1991, "Realism, Anti-Foundationalism, and the Enthusiasm for Natural Kinds", *Philosophical Studies*, vol. 61, pp. 127–148.
- Butler, K., 1994, "Neural Constraints in Cognitive Science", *Minds and Machines*, vol. 4, pp. 129–162.
- Churchland, P.M., 1989, *A Neurocomputational Perspective*, MIT Press, Cambridge, Mass.
- Churchland, P.S., 1986, *Neurophilosophy*, MIT Press, Cambridge, Mass.
- Clark, A., 1993, *Associative Engines*, MIT Press, Cambridge, Mass.
- , 1989, *Microcognition*, MIT Press, Cambridge, Mass.
- Corbí, J.E., 1993, "Classical and Connectionist Models: Levels of Description", *Synthese*, vol. 95, pp. 141–168.
- Darden, L. and N. Maull, 1977, "Interfield Theories", *Philosophy of Science*, vol. 43, pp. 44–64.
- Davidson, D., 1993, "Thinking Causes", in J. Heil and A. Mele (eds.), *Mental Causation*, Clarendon, Oxford.
- , 1974, "Psychology as Philosophy", in S.C. Brown (ed.), *Philosophy of Psychology*, Macmillan, London.
- , 1973, "Material Mind", in P. Suppes *et al.* (eds.), *Logic, Methodology and Philosophy of Science, IV*, North-Holland, Amsterdam.
- , 1970, "Mental Events", in L. Foster and J. Swanson (eds.), *Experience and Theory*, Duckworth, London.
- Davies, M. and T. Stone (eds.), 1995, *Folk Psychology*, Blackwell, Oxford.
- Dennett, D.C., 1987, *The Intentional Stance*, MIT Press, Cambridge, Mass.

- Fodor, J., 1987, *Psychosemantics. The Problem of Meaning in the Philosophy of Mind*, MIT Press, Cambridge, Mass.
- , 1983, *The Modularity of Mind*, MIT Press, Cambridge, Mass.
- , 1974, "Special Sciences: or the Disunity of Sciences as a Working Hypothesis", *Synthese*, vol. 28, pp. 77–115.
- Fodor, J.A. and B.P. McLaughlin, 1990, "Connectionism and the Problem of Systematicity: Why Smolensky's Solution Doesn't Work", *Cognition*, vol. 35, pp. 183–204. (Reprinted in Horgan and Tienson 1991.)
- Fodor, J.A. and Z.W. Pylyshyn, 1988, "Connectionism and Cognitive Architecture: A Critical Analysis", *Cognition*, vol. 28, pp. 3–71.
- Foster, C.L., 1992, *Algorithms, Abstraction and Implementation. Levels of Detail in Cognitive Science*, Academic Press, San Diego.
- Gardner, H., 1985, *The Mind's New Science*, Basic Books, New York.
- Goodman, N., 1965, *Fact, Fiction and Forecast*, Bobbs-Merrill, Indianapolis.
- Hardcastle, V.G., 1997, "Distinctions without Differences: Commentary on Horgan and Tienson's Connectionism and the Philosophy of Psychology", *Philosophical Psychology*, vol. 3, pp. 373–384.
- , 1996, *How to Build A Theory in Cognitive Science*, State University of New York Press, New York.
- Hooker, C.A., 1981, "Towards a General Theory of Reduction. Part. I: Historical and Scientific Setting; Part. II Identity and Reduction; Part III: Cross Categorical Reduction", *Dialogue*, vol. 20, pp. 38–59, 201–236, 496–529.
- Horgan, T. and J. Tienson, 1997, "Précis of *Connectionism and the Philosophy of Psychology*", *Philosophical Psychology*, vol. 10, pp. 337–354.
- , 1996, *Connectionism and the Philosophy of Psychology*, MIT Press, Cambridge, Mass.
- , 1994, "A Nonclassical Framework for Cognitive Science", *Synthese*, vol. 101, pp. 305–345.
- , 1993, "Levels of Description in Nonclassical Cognitive Science", *Philosophy*, Supplement, vol. 34, pp. 159–188.
- , 1992, "Cognitive Systems as Dynamical Systems", *Topoi*, vol. 11, pp. 27–44.
- (eds.), 1991, *Connectionism and the Philosophy of Mind*, Kluwer, Dordrecht.
- , 1989, "Representations without Rules", *Philosophical Topics*, vol. 17, pp. 147–174.

- Kitcher, P., 1981, "Explanatory Unification", *Philosophy of Science*, vol. 48, pp. 507–531.
- Lakoff, G., 1987, *Women, Fire, and Dangerous Things: What Categories Reveal about the Mind*, University of Chicago Press, Chicago.
- Laudan, L., 1977, *Progress and Its Problems. Towards a Theory of Scientific Growth*, University of California Press, Berkeley.
- Marr, D., 1982, *Vision*, Freeman, San Francisco.
- McCauley, R.N., 1995, "Explanatory Pluralism and the Co-evolution of Theories in Science", in R.N. McCauley (ed.), *The Churchlands and their Critics*, Blackwell, Oxford.
- , 1986, "Intertheoretic Relations and the Future of Psychology", *Philosophy of Science*, vol. 56, pp. 179–199.
- Medin, D.L. and M.M. Schaffer, (1978, "A Context Theory of Classification Learning", *Psychological Review*, vol. 85, pp. 207–238.
- Millikan, R.G., 1999, "Historical Kinds and the 'Special Sciences'", *Philosophical Studies*, vol. 95, pp. 45–65.
- Nagel, E., 1961, *The Structure of Science*, Harcourt, Brace and World, New York.
- Newell, A., 1980, "Physical Symbol Systems", *Cognitive Science*, vol. 4, pp. 135–183.
- , 1990, *Unified Theories of Cognition*, Harvard University Press, Cambridge, Mass.
- , 1982, "The Knowledge Level", *Artificial Intelligence*, vol. 18, pp. 87–127.
- Newell, A. and H.A. Simon, 1976, "Computer Science as Empirical Enquiry: Symbols and Search", *Proceedings of the Association for the Computing Machinery*, vol. 19, pp. 113–126. (Reprinted in J. Haugeland (ed.), 1981, *Mind Design*, MIT Press, Cambridge, Mass., pp. 35–66.)
- Pylyshyn, Z.W. (ed.), 1987, *The Robot's Dilemma*, Ablex, Norwood.
- , 1984, *Computation and Cognition*, MIT Press, Cambridge, Mass.
- Ramsey, W., S.P. Stich and D.E. Rumelhart (eds.), 1991, *Philosophy and Connectionist Theory*, Lawrence Erlbaum, Hillsdale.
- Richardson, R.C., 1999, "Cognitive Science and Neuroscience: New Wave Reductionism", *Philosophical Psychology*, vol. 12, pp. 297–307.
- Rivière, A., 1991, *Objetos con mente*, Alianza, Madrid.

- Rosch, E., 1978, "Principles of Categorization", in E. Rosch and B. Lloyd (eds.), *Categorization and Cognition*, Lawrence Erlbaum, Hillsdale.
- Rosenberg, J.F., 1994, "Comments on Bechtel, 'Levels of Description and Explanation in Cognitive Science'", *Minds and Machines*, vol. 4, pp. 27–37.
- Schaffner, K.F., 1967, "Approaches to Reduction", *Philosophy of Science*, vol. 34, pp. 137–147.
- Schouten, M.K.D. and H. Looren de Jong, 1999, "Reduction, Elimination, and Levels: The Case of the LTP-learning Link", *Philosophical Psychology*, vol. 12, pp. 237–262.
- Silvers, S., 1996, "Rational Reconstruction and Immature Science", *Philosophical Psychology*, vol. 9, pp. 93–109.
- Smolensky, P., 1991, "The Constituent Structure of Connectionist Mental States: A Reply to Fodor and Pylyshyn", in Horgan and Tienson 1991, pp. 281–308. (Reprinted from *The Southern Journal of Philosophy*, vol. 26 (suppl.), pp. 137–163.)
- , 1988, "On the Proper Treatment of Connectionism", *Behavioral and Brain Sciences*, vol. 11, pp. 1–23.
- Smolensky, P., G. Legendre and Y. Miyata, 1993, "Integrating Connectionist and Symbolic Computation for the Theory of Language", *Current Science* vol. 64, pp. 381–391.
- Sterelny, K., 1990, *The Representational Theory of Mind*, Blackwell, Oxford.
- Stich, S., 1983, *From Folk Psychology to Cognitive Science*, MIT Press Cambridge, Mass.
- Suppes, P., 1957, *Introduction to Logic*, Van Nostrand, New York.
- Tienson, J., 1997, "What the Differences Are: Reply to Hardcastle", *Philosophical Psychology*, vol. 3, pp. 385–389.
- Von Eckardt, B., 1993, *What Is Cognitive Science?*, MIT Press, Cambridge, Mass.

*Received January 6, 2003; revised July 1, 2004; accepted July 7, 2004.*