# Explanatory Pluralism And The Co-Evolution Of Theories In Science *

**Robert N. McCauley**

## 1. Introduction

Over the past decade or so Patricia and Paul Churchland have made major contributions to philosophical treatments of intertheoretic reduction in science. The historic importance of this issue in the philosophy of science is patent and so, therefore, is the importance of the Churchlands' contributions. Their insistence on the centrality of this issue to discussions in the philosophy of mind may, however, be even more praiseworthy in an era when many in that field (even among those who claim the mantle of naturalism) make repeated declarations about the status of the pertinent sciences and the mind-body problem generally in what often appears to be blithe ignorance of both those sciences and the relevant literature in the philosophy of science since 1975.

In a recent, joint paper the Churchlands (1990) discuss and largely defuse five well-worn objections (concerning qualia, intentionality, complexity, freedom, and multiple instantiation) to the reduction of psychology to neurobiology. My concerns with that putative reduction and with the Churchlands' account of the overall process are of a very different sort.

Two models have traditionally dominated discussions of intertheoretic relations. After briefly surveying the contrasts between them, section 2 examines how the Churchlands' account of these relations in terms of a continuum of intertheoretic commensurability captures those models' respective advantages in a single proposal. That section ends by examining how Patricia Churchland's subsequent discussions of the co-evolution of theories enhances this account by exploring some of its underlying dynamics. In short, the co-evolution of theories concerns cross-scientific interactions that change the position of a particular intertheoretic relationship on the Churchlands' continuum.

In section 3 I locate some revealing equivocations in the Churchlands' discussions of "the co-evolution of theories" by distinguishing three possible interpretations of that notion that wind their ways through the Churchlands' work and through *Neurophilosophy* in particular. With the aid of a distinction concerning levels of analysis that I have developed elsewhere, I argue, in effect, that the Churchlands' account of the co-evolution of theories and their model of intertheoretic reduction obscure critical distinctions between three quite different types of intertheoretic relations. Section 4 positions these three types within a more fine-grained account of intertheoretic relations that will offer a basis for evaluating their relative merits as analyses of the interface of psychology and neuroscience.

One of these three, the picture of co-evolution modeled on the dynamics of scientific revolutions, has attracted the most attention. This interpretation has encouraged the recurring eliminativist inclinations concerning folk psychology for which the Churchlands are renown, but, of the three, it is also the interpretation that is least plausible as an analysis of the relations between psychology and neuroscience. Psychology (folk or otherwise) may well undergo substantial revision, and future scientific progress may well lead to the elimination of some psychological theories, but the Churchlands have offered an unhelpfully oversimplified account of the intertheoretic dynamics in question.

In section 5 I shall support and elaborate upon another of the interpretations of co-evolution that emerges from *Neurophilosophy* by, among other things, examining a case (concerning the connectionist network, NETtalk) that the Churchlands and their collaborators have highlighted. This third interpretation recognizes not merely the value of integrating scientific disciplines but of preserving a plurality of semi-autonomous explanatory perspectives. Although the Churchlands now often seem to favor this third interpretation too, some of their comments continue to conflate the three distinct types of intertheoretic relations.

## 2. Three Philosophical Models of Intertheoretic Relations in Science

Until the late 1970s (at least) two models of intertheoretic relations in science dominated philosophers' attentions. The first, a general purpose model of intertheoretic relations, was deeply rooted in logical empiricism; the second, in effect a model of theory change, emerged largely in reaction to the first (Bechtel 1986). I shall briefly discuss them in order.

Although Ernest Nagel's *The Structure of Science* (1961) contains the most time-honored treatment of theory reduction, Robert Causey's *Unity of Science* (1977) probably provides the most comprehensive discussion of the topic. Their general approach to theory reduction proceeds within the constellation of commitments that characterize logical empiricism, including the assumptions that a satisfactory account of scientific rationality requires heed to justificatory considerations only, that scientific theories are best understood as complex propositional structures and best represented via formal reconstructions, that scientific explanation results from the deduction of explananda from scientific laws, that scientific progress results from the subsumption of reigning theories by theories of even greater generality, and that science ultimately enjoys an underlying unity of theory and ontology.

This model conceives theory reduction as a special case of deductive-nomological explanation. It is a special case because the explanandum is not a statement describing some event but rather a law of the reduced theory. In order to carry out such reductions, the premises in the most complex cases of heterogeneous reductive explanations must include
(1) at least one law from the reducing theory:
(2) statements indicating the satisfaction of the requisite initial conditions specified in that law;
(3) bridge laws which systematically relate–*within a particular domain* delineated by appropriate boundary conditions–the terms from the pertinent law(s) of the reducing theory to those from the law of the reduced theory;
(4) statements indicating the satisfaction of those boundary conditions (under which the events described in the law of the reducing theory realize the events described in the law of the reduced theory that is to be explained).
Such premises permit a straightforward deduction of the law of the reduced theory.

Because the boundary conditions included in the bridge laws are cast in terms of predicates characteristic of the reducing theory, the reduction reflects an asymmetry between the two theories. The reducing theory explains the reduced theory, finally, because the reducing theory encompasses a wider array of events within its explanatory purview. This set of events, presumably, includes all of the events the reduced theory explains and more, so that the principles of the reducing theory are both more general and more fundamental. The most popular showcase illustration is the reduction of the laws of classical thermodynamics to the principles of statistical mechanics.

When the reducing theory operates at a lower level of analysis than the reduced theory, the added generality of its principles is a direct function of this fact. These are cases of *microreductions* where a lower level theory and its ontology reduce a higher level theory and its ontology (Oppenheim and Putnam 1958, Causey 1977). Microreductionists hold that if we can exhaustively describe and predict upper level (or macro) entities, properties, and principles in terms of lower level (or micro) entities, properties, and principles, then we can reduce the former to the latter and replace, at least in principle, the upper level theory.

Virtually all discussions of intertheoretic relations presuppose this arrangement among (and within) the sciences in terms of levels of analysis. (See, for example, Churchland and Sejnowski 1992, pp. 10-11.) Numerous considerations contribute to the depiction of the architecture of science as a layered edifice of analytical levels (Wimsatt 1976). Ideally, moving toward lower levels involves moving toward the study of increasingly simple systems and entities that are ubiquitous, enduring, and small. Conversely, moving from lower to higher level sciences involves moving toward studies of larger, rarer systems of greater complexity and (often) less stability and whose history is less ancient. Because the altitude of a level of analysis is directly proportional to the complexity of the systems it treats, higher level sciences deal with increasingly restricted ranges of events having to do with increasingly organized physical systems.[1]

As a simple matter of fact, often more than one configuration of lower level entities can realize various higher level kinds (especially when functionally characterized). The resulting multiple instantiations highlight both the importance and the complexity of the boundary conditions in the bridge laws of heterogeneous microreductions. Critics of the microreductionist program (e.g., Fodor 1975) see that complexity as sufficient grounds for questioning the program's feasibility in the case of the special sciences, while more sympathetic participants in these discussions such as Robert Richardson (1979) and the Churchlands (1990) suggest that when scientists trace out such connections between higher and lower level entities in specific domains they vindicate the overall strategy, while recognizing the *domain specificity* of its results.

Reductionists differ among themselves as to the precise connections between entities at different levels that are required for successful reductive explanation. They all agree, however, that the theories which are parties to the reduction should map on to one another well enough to support systematic connections, usually contingent identities, between some, if not all, of the entities that populate them. The test of the resulting contingent identities is met, ultimately, by the explanatory successes the reductions accomplish (McCauley 1981; Enc 1983).

Feyerabend (1962) and Kuhn (1970) are the most prominent proponents of the second major account of intertheoretic relations. They forged their early discussions largely in response to both the logical empiricist program and its reductionist blueprint for scientific progress. Feyerabend emphasized how scrutiny of many of the showcase illustrations of intertheoretic reductions revealed the *failure* of these cases to conform to the logical empiricists' model. Kuhn discussed numerous examples in the history of science where successive theories were not even remotely plausible candidates for the sort of smooth transitions the standard reductive model envisions. Instead, Kuhn proposed that progress in science consists of extended intervals of relative theoretical stability punctuated by periodic revolutionary upheavals. Both hold that the cases in question involve conflicts between *incommensurable* theories.

Although the subsequent literature is rife with assessments of this claim (Thagard 1992 offers the most suggestive of recent treatments), the critical point for now is that, whatever incommensurability amounts to, it stands in stark opposition to any model of intertheoretic

relations that requires neat mappings between theories' principles and ontologies capable of supporting strict deductive-nomological explanations. The history of science provides ample evidence that where such incompatibility is sufficiently severe the theory and its ontology that are eventually deemed deficient undergo elimination. Stahl's system of chemistry is the preferred illustration, but Darwin's theory of inheritance could serve just as well.

The unmistakable sense that both of these models of intertheoretic relations describe some actual cases fairly accurately and that they each capture important insights about the issues at stake, their profound conflicts notwithstanding, could induce puzzlement. An account of intertheoretic relations in terms of a continuum of commensurability that Paul Churchland (1979) initially sketched and which the Churchlands have subsequently developed (P.S. Churchland 1986, pp. 281f.; Churchland and Churchland 1990) substantially resolves that perplexity by reconciling those conflicts and allotting to each model a measure of descriptive force.

The Churchlands point out that, in fact, different cases of intertheoretic relations vary considerably with respect to the commensurability of the theories involved. So, they propose that such cases fall along a continuum of relative intertheoretic commensurability, where, in effect, the two models sketched above constitute that continuum's end-points.

One end of the continuum represents cases where intertheoretic mapping is extremely low or even absent. These are cases of *radical* incommensurability where revolutionary science and the complete elimination of inferior theories ensue. Whatever vagueness may surround the notion of 'incommensurability,' the Churchlands are clearly confident that the developments which brought about the elimination of the bodily humours, the luminiferous ether, caloric fluid, and the like, involve sufficiently drastic changes to justify the sort of extreme departures from the traditional model of reduction that Kuhn and Feyerabend advocated.

At the other end of this continuum, where the mapping of one theory on another is nearly exhaustive and the former theory's ontology is composed from the entities the latter theory countenances, the most rigorous models of theory reduction most nearly apply (e.g., Causey 1972). The constraints proponents have imposed on theory reduction are so demanding that it is a fair question whether *any* actual scientific case qualifies. The Churchlands have urged considerable relaxation of the conditions necessary for intertheoretic reduction. Instead of conformity to the rigorous logical and ontological constraints traditional models impose, Paul Churchland (1979; see too Hooker 1981; Bickle 1992) suggests that the reducing theory need only preserve an "equipotent image" of the reduced theory's most central explanatory principles. The reduction involves an *image*, since the reducing theory need not duplicate every feature of the reduced theory's principles, but only enough of their salient ones to suggest their general character and to indicate their systematic import (see Schaffner 1967). That image is *equipotent*, though, since the reducing theory's principles will possess all of the explanatory and predictive power of the reduced theory's principles–and more. From the standpoint of traditional models, Churchland proposes a form of *approximate* reduction, which falls well short of the logical empiricists' standards, but which also suggests how true theories (e.g., the mechanics of relativity) can correct and even approximately reduce theories that are false (e.g., classical mechanics). Switching to the metaphor of imagery is appropriate, since, as William Wimsatt (1976, p. 218) noted over a decade ago, if the standard models of reduction allege that a false theory follows from a true one, the putative deduction had better involve an equivocation somewhere!

In recent years the Churchlands have each enlarged on this continuum model. For example, within his neurocomputational program Paul Churchland has advanced a prototype activation model of

explanatory understanding that, presumably, includes the understanding that arises from *reductive* explanations.

Churchland holds that the neurocomputational basis of explanatory understanding resides in the activation of a prototype vector within a neural network in response to impinging circumstances. A distributed representation of the prototype in the neural network constitutes the brain's current best stab at detecting an underlying pattern in the blooming, buzzing confusion. For Churchland explanatory understanding *is* an array of inputs leading to the activation of one of these existing prototypes as opposed to another.

Churchland insists that the activation of a prototype vector increases, rather than diminishes, available information. It involves a "speculative *gain*" in information (1989, p. 212). Thus, contrary to anti-reductionist caricature, this account of explanatory understanding implies that reductive explanations *amplify* our knowledge. The originality of the insights a reductive explanation offers depends upon the novel application of existing cognitive resources, i.e., of an individual's repertoire of prototype vectors. Consequently, reductive explanation involves neither the generation of new schemes nor the destruction of old ones. The approximate character of intertheoretic reductions is a function of this "conceptual redeployment" on which they turn (1989, p. 237). In conceptual redeployment a developed conceptual framework from one domain is enlisted for understanding another. In short, successful reductive explanation rests on an analogical inference by virtue of which we deem an image of a theory equipotent to the original. Having established the initial applicability of an existing, alternative prototype vector, it inevitably undergoes a reshaping as a consequence of exposure to the newly adopted training set. This reshaping of activation space is the neurocomputational process that drives the remaining co-evolution of the reductively related theories.

In her discussions of the co-evolution of theories, Patricia Churchland has introduced a dynamic element into the continuum model. She suggests that the position of two theories' relations on this continuum can change over time as they each undergo adjustments in the light of one another's progress.

The suggestion that scientific theories co-evolve arises from an analogy with the co-evolution of species and from the picture of the sciences briefly outlined above. On the co-evolutionary picture the sciences exert selection pressures on one another in virtue of a general concern for supplying as much coherence as possible among our explanatory schemes. If the various sciences are arranged in tiers of analytical levels, then each will stand at varying distances from the others in this structure. Typically, proximity is a central consideration in assessing the force of selection pressures. Thus, the pivotal relationships are those between a science and those sciences at immediately adjacent levels. For example, the presumption is that the neurosciences below and the socio-cultural sciences above are more likely to influence psychology than are the physical sciences, since they are located below the neurosciences and, therefore, at an even greater distance.

It is this process of the co-evolution of theories and the Churchlands' account of it that will dominate the remainder of this paper. I shall attend to its implications for the relationship of cognitive psychology to the sort of neurocomputational modeling that the Churchlands endorse.

### 3. Three Ways Theories Might Co-evolve

Patricia Churchland's *Neurophilosophy* (1986) contains the most extensive discussion of reduction in terms of the co-evolution of theories available.[2] Churchland focuses on the relation between

neuroscience and psychology, but her discussion clearly aspires to morals that are general. Her comments at various points seem to support three different co-evolutionary scenarios, though two of them are, quite clearly, closely related. The three are distinguished by the locations on the Churchlands' continuum to which they predict co-evolving theories will incline.

On some occasions Churchland suggests that psychology and the neurosciences will co-evolve in the direction of approximate reduction. She states, for example, that "the co-evolutionary development of neuroscience and psychology means that establishing points of reductive contact is more or less inevitable. . . . The heart of the matter is that if there is theoretical give and take, then the two sciences will knit themselves into one another" (1986, p. 374). The metaphor of two sciences knitted into one another implies an integration that is tight, orderly, and detailed. Although Churchland, presumably, does not think that that integration will satisfy the traditional microreductionists' stringent demands on intertheoretic mapping, talk of knitting two sciences into one another, the on-going pursuit of a unified model of *reduction* (Churchland and Churchland 1990), and a new interest in establishing psycho-physical identities echo commitments of traditional microreductionism, where the sort of reductive contact in question led to talk of an "in principle replaceability" of the reduced theory in which the lower level theory enjoys both explanatory and metaphysical priority. More recently, the Churchlands have been clear about the futility of attempts to replace upper level theories, but they still generally subscribe to the explanatory and metaphysical priority of the lower level theory–especially in the case of psychology and neuroscience. (See, for example, P.S. Churchland 1986, pp. 277, 294, and 382.) Certainly, a co-evolutionary account of intertheoretic relations has no problem translating the general microreductive *impulse*. Within this framework it amounts to the claim that the selection pressures that the science at the level of analysis *below* that of the theory in question exerts will have an overwhelmingly greater effect on that theory's eventual shape and fate than will the sciences above (see section 5).

On the Churchlands' account, such intertheoretic integration would enable the neurosciences to supply an equipotent image of psychological principles. Paul Churchland's speculations about the neural representation of the sensory qualia associated with color vision might constitute an appropriate illustration. The fit between our common sense notions about our experiences of colors and the system of neural representation he proposes is quite neat (1989, p. 102-08). Hereafter I shall refer to this sense of the co-evolution of theories as "co-evolution$_M$," i.e., co-evolution in the direction of approximate microreduction. In the Churchlands' joint discussion (1990, chapter 6.1), where it plays both a predictive and normative role, this notion of reduction receives considerable attention. The Churchlands clearly hold that "it is reasonable to expect, and to work toward, a reduction of all psychological phenomena to neurobiological and neurocomputational phenomena" (1990, p. 249).

Co-evolution$_M$ is not the only account of co-evolution in *Neurophilosophy*, for, as the Churchlands have subsequently asserted, in the case of psychology and neuroscience, "there are conflicting indications" about the direction in which conjectures at these two levels of analysis will likely co-evolve (1990, p. 253). If integration is the fate of psychology and neuroscience, Patricia Churchland repeatedly hints that this will only occur *after* psychology's initial demolition and subsequent reconstruction in accord with the mandates of the neurosciences. She claims, for example, that " . . . the possibility that psychological categories will not map one to one onto neurobiological categories . . . does not look like an obstacle to reduction so much as it predicts a fragmentation and reconfiguration of the psychological categories" (1986, p. 365). With this second view, as with the first, no question arises about where the blame lies, if the theories of psychology and neuroscience fail to map onto one another neatly. (See Wimsatt 1976.) *At least for the short term*, Churchland

seems to expect that this intertheoretic relation will migrate in just the *opposite* direction on the continuum of intertheoretic commensurability from what co-evolution$_M$ predicts, i.e., toward a growing *in*commensurability that predicts a fragmentation of *psychological* categories.

If the "fragmentation and reconfiguration" of psychological categories involved only the elaboration or adjustment (or even the in principle replaceability) of psychological theories by discoveries in the neurosciences, co-evolution$_M$ might suffice. On this second view, though, this process can lead to the eventual eradication of major parts of psychology. So, for example, Churchland remarks that "there is a tendency to assume that the capacities at the cognitive level are well defined . . . in the case of memory and learning, however, the categorial definition is far from optimal, and *remembering stands to go the way of impetus*" (1986, p. 373, emphasis added[3]). Here Churchland anticipates that just as the new physics of Galileo and his successors ousted the late medieval theory of impetus, so too shall advances in neuroscience dispose of psychologists' speculations about memory. This, then, is co-evolution$_S$ (co-evolution producing the eliminations of theories characteristic of *s*cientific revolutions) in which the theoretical perspectives of two neighboring sciences are so disparate that eventually the theoretical commitments of one must go--in the face of the other's success.

Co-evolution$_S$ underlies the position for which the Churchlands' advocacy has been famous, viz., eliminative materialism.[4] They have contended that progress in the neurosciences will probably bring about the elimination of folk psychology as well as any other psychological theories that involve commitments to the propositional attitudes (presumably, including much of mainstream cognitive and social psychology). Just as scientists banished phlogiston and caloric fluid, so too will the propositional attitudes be expelled as neuroscience progresses. The psychological conjectures in question (will) fail to match the descriptive, explanatory, and predictive successes of their neuroscientific competitors. Moreover, their substantial dissimilarities to those alleged competitors preclude any sort of reconciliation. Consequently, numerous theoretical notions in psychology stand to go the way of impetus. This is the predicted result when the Churchlands emphasize, among those "conflicting indications," the *uncongenial* relations between psychology and neuroscience.

Revising their extreme eliminativism, the Churchlands sometimes seem to intend these two interpretations to address different stages in the co-evolutionary process (as I suggested above): first, the demolition of much current psychology via co-evolution$_S$ followed by the reconstruction of a neuroscientifically inspired psychology via co-evolution$_M$. The crucial point for now is that these two interpretations of co-evolution hold that the relationship between two theories will, over time, shift in one direction (as opposed to the other) on the Churchlands' continuum.

An obvious question arises, though. If either direction is possible, then what are the variables that determine the direction of any shift? (This question presses the revised version of eliminativism no less than the original.) The Churchlands have not addressed this question directly, because they have recognized that the complexities of the intertheoretic relations in question and of the relationship of psychology and neuroscience, in particular, require more. Enter the third interpretation.

One of Patricia Churchland's extended comments about her general model of reduction (1986, pp. 296-7) is especially revealing, since it reflects at various points the influence of all three interpretations.

> . . . some misgivings may linger about the possibility of reduction should it be assumed that a reductive strategy means an exclusively bottom-up strategy . . .

> These misgivings are really just bugbears, and they have no place in my framework for reduction.
>
> . . . if the reduction is smooth, its reduction gives it [the reduced theory]– and its phenomena–a firmer place in the larger scheme . . . If the reduction involves a major correction, the corrected, reduced theory continues to play a role in prediction and explanation . . . Only if one theory is eliminated by another does it fall by the wayside.
>
> . . . coevolution . . . is certain to be more productive than an isolated bottom-up strategy.

The second paragraph traces points on the continuum. It alludes initially to co-evolution$_M$–its final sentence to co-evolution$_S$. It is the first and third paragraphs, though, where shadows of a third interpretation appear.

Closely related to co-evolution$_M$ is co-evolution$_P$ (co-evolution as explanatory pluralism). Their many similarities notwithstanding, it is worth teasing them apart. As a first pass, where co-evolution$_M$ anticipates increasing intertheoretic integration largely guided by and with a default preference for the lower level, co-evolution$_P$ construes the process as preserving a diverse set of partially integrated yet semi-autonomous explanatory perspectives–where that non-negligible measure of analytical independence rests at each analytical level on the explanatory success and the epistemic integrity of the theories and on the suggestiveness of the empirical findings. Co-evolution$_M$, in effect, holds that selection pressures are exerted exclusively from the bottom up, whereas co-evolution$_P$ attends to the constraints imposed by the needs and demands of theories operating at higher levels.

These apparently small differences are but the fringe skirmishes of some of the most basic epistemological and metaphysical battles in the philosophy of science. Space limitations preclude extensive development, but broadly, if they are not persuaded by co-evolution$_S$, physicalists prefer co-evolution$_M$, since it suggests a science unified in both theory and ontology that accords priority to the lower (i.e., physical) levels. More pragmatically minded philosophers opt for co-evolution$_P$, foregoing assurances of and worries about a unified science and metaphysical purity in favor of enhanced explanatory resources. For nearly a decade now the Churchlands have been negotiating their interests in unified science and metaphysical purity on the one hand with their interests in enhanced explanatory resources and internalism on the other. (See McCauley 1993 and note 10 below.) The relaxation of their eliminativism and their emerging preference for co-evolution$_P$ indicate the influence of pragmatic currents in their thought.

Co-evolution$_P$ is prominent in *Neurophilosophy* and even more so since.[5] Patricia Churchland claims that " . . . the history of science reveals that co-evolution of theories has typically been mutually enriching," that "[r]esearch influences go up and down and all over the map," that "co-evolution typically is . . . interactive . . . and involves one theory's being susceptible to correction and reconceptualization at the behest of the cohort theory," and that "psychology and neuroscience should each be vulnerable to disconfirmation and revision at any level by the discoveries of the other" (1986, pp. 363, 368, 373, and 376).

Figure 6.2.1 seems the most plausible interpretation of the relationship between these three notions of co-evolution and the earlier continuum model; it roughly indicates the regions of that continuum where the cases covered by the three types of co-evolution end up. (See Churchland and Churchland 1990, p. 252.)

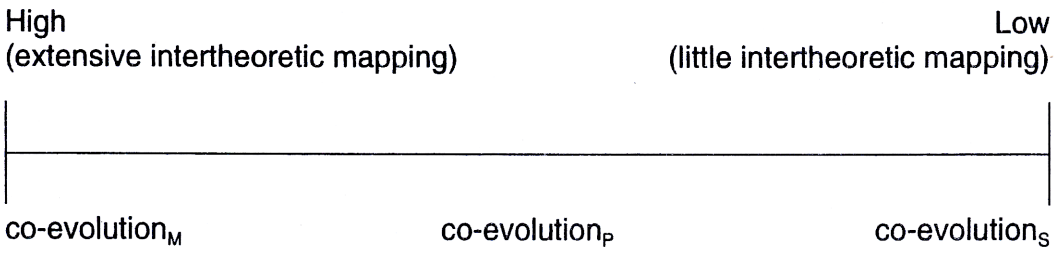


Figure 6.2.1. Three notions of co-evolution situated on the Churchlands' continuum.

Section 4 will suggest that the picture in Figure 6.2.1 of intertheoretic relations and of co-evolution, in particular, is oversimplified to the point of distortion. The intertheoretic dynamics of scientific revolutions are quite different from those of approximate microreduction and explanatory pluralism. Crucially, co-evolution$_P$ is incompatible with co-evolution$_S$. The *mutual* intertheoretic enrichment co-evolution$_P$ envisions will not arise, if neuroscience is radically reconfiguring (let alone eliminating) psychology. Neither the history of science nor pragmatic accounts of scientific practice offer much reason to think that co-evolution$_S$ provides either an accurate description or a useful norm for the relationship between psychology and neuroscience or for any such relationship between theories in sciences operating at *different* analytical levels.

The differences between co-evolution$_P$'s and co-evolution$_M$ are also important. At stake is the question of the relative priority of neuroscientific (lower level) and psychological (upper level) contributions to the science of the mind/brain. This topic will dominate section 5. In criticizing co-evolution$_S$ and curtailing co-evolution$_M$, the aim of the next two sections is, ultimately, to endorse and develop the notion of explanatory pluralism.

**4. Exploring Explanatory Pluralism: Debunking Co-evolution$_S$**

Enlisting a distinction Wimsatt (1976) introduced between *intralevel* and *interlevel* contexts, I have previously developed a model of intertheoretic relations that discloses why we should not expect advances in neuroscience to eliminate much psychology directly (McCauley 1986). More generally, it suggests that co-evolution$_S$ does not very happily model the co-evolving relations of theories at different levels.

The sorts of unequivocal eliminations of theories and ontologies that co-evolution$_S$ countenances arise in *intra*level contexts involving considerable incommensurability. These contexts concern changes *within* a particular science over time. They include the classic cases that philosophers group under the rubric of "scientific revolutions" – impetus, phlogiston, caloric fluid, and the like. Within a particular level of analysis some newly proposed theory proves superior to its immediate predecessor with which it is substantially discontinuous. When the scientific community opts for this new theory, most traces of its predecessor rapidly disappear. Since they offer incompatible accounts of many of the same phenomena, the new theory *explains* the old theory *away*.

By contrast, intralevel situations where the mappings between theories are reasonably good fall near the other end of the Churchlands' continuum. Here the new theory *explains* its predecessor which it also typically corrects. Scientists regard the earlier theory's domain as a special case to which the new theory applies and for which the old theory continues to suffice as a useful

calculating heuristic. Although corrected and incorporated as a special case into a more general theory, Newton's laws of motion work well for most practical purposes.

A new theory disrupts science less to the extent it preserves (rather than overthrows) the cherished insights and conceptual apparatus of its predecessors. It may require reinterpretation of established notions ("planets," "genes," "grammar acquisition," etc.), but changes are evolutionary only when they preserve a fair measure of intensional and extensional overlap with their predecessors. When succeeding theories in some science are largely continuous, no one speaks of elimination. The change is evolutionary, not revolutionary. Consequently, the new theory is perfectly capable of providing an equipotent image of the old. These are the cases where the new theory overwhelmingly *inherits* the evidence for the old.

Revolutionary or evolutionary, progress within some science eliminates features of earlier theories eventually. In revolutionary settings the changes are abrupt and the elimination is (relatively) immediate. In evolutionary contexts incompatibilities accrue over time. Although the transition from one theory to its immediate successor may be more or less smooth, over a series of such transitions all traces of ancestral theories may completely disappear. Consider the fate of "natural motions" from Aristotelian through Newtonian mechanics (McCauley 1986, pp. 192-93). Similarly, over the past hundred years the "memory trace" has undergone considerable evolutionary transformation. Some theorists would argue that the reinterpretations have been so substantial that the original notion (and what it allegedly referred to) has virtually vanished.[6]

*Inter*level relations concern theories at *different* (typically neighboring) levels of analysis at a particular point in time (in contrast to intralevel cases concerned with successive theories at the *same* level of analysis). The Churchlands' continuum maps onto interlevel cases too.

When sciences at adjoining levels enjoy substantial intertheoretic mapping (in situations approximating classic microreductions) they heavily constrain one another's form–otherwise, why would anyone have attempted to characterize their relations in terms of deductive logic and identity statements? This is the effect of the knitting of two sciences into one another that co-evolution$_M$ envisions. A well integrated lower level theory has resources sufficient to reproduce the explanatory and predictive accomplishments of the corresponding upper level theory, however, this often comes at considerable computational expense. As the Churchlands have emphasized, this does not disgrace the higher level theory nor lead to the evaporation of the phenomena it seeks to explain.

When considering interlevel cases with relatively unproblematic intertheoretic relations, the Churchlands, like the traditional reductionists before them[7], have focused exclusively on their *resemblances* to the intralevel settings described above. (See, for example, P.S. Churchland 1986, p. 294.) After all, here too the elaborations of the upper level theory's central concepts that the lower level, *reducing* theory offers often correct the less fine-grained, upper level theory's pronouncements. However, because the theories are tightly knit, the upper level theory still provides a useful and efficient approximation of the lower level theory's results. This sounds quite like the cases of scientific evolution described above.

Beneath these resemblances, though, lie small but revealing differences. First, unlike the evolutionary intralevel cases, the reduced theory in interlevel situations does not stand in need of technical correction in *every* case. For a few situations at least, its results will conform precisely with those of the lower level theory, because for these cases it adequately summarizes the effects of all relevant lower level variables.[8] This contrasts with the *inescapable*, if often negligible (from a

practical standpoint), divergence of the calculations of some theory and its successor, such as classical mechanics and the mechanics of relativity. (See Churchland and Churchland 1990, p. 251.) In interlevel cases corrections can arise because the upper level theory is insufficiently fine-grained to handle certain problems. By contrast, in intralevel cases corrections always arise because the earlier theory is wrong–by a little in evolutionary cases, by a lot in revolutionary ones. It follows that the upper level theory is not always a *mere* calculating heuristic (as the replaced predecessor is in cases of scientific evolution). Moreover, the upper level theory's heuristic advantages in well integrated interlevel contexts are typically *enormous*, compared with intralevel cases. The divergence of computational effort between the classical and statistical solutions for simple problems about gases (an interlevel case) dwarfs that between classical mechanics and the mechanics of relativity for simple problems about motion (an intralevel case). Of a piece with this observation, the Churchlands quite accurately describe the quantum calculations of various chemical properties (another interlevel case) as "daunting" (1990, p. 251). Finally, the upper level theory lays out regularities about a subset of the phenomena that the lower level theory encompasses but for which it has neither the resources nor the motivation to highlight. That is the price of the lower level theory's generality and finer grain.

If these considerations are not compelling, scrutiny of interlevel circumstances that support relatively little intertheoretic mapping reveals far more important grounds for stressing the distinction between interlevel and intralevel settings. Here two sciences at adjacent levels address some common explananda under different descriptions, but their explanatory stories are largely (though not wholly) incompatible. On the Churchlands' view, this is just the relationship between neuroscience and most of folk psychology, and if remembering is to go the way of impetus, the relationship between neuroscience and some important parts of scientific psychology as well.

If all of these intertheoretic relations should receive a *unified* treatment, as traditional reductionists, the Churchlands (e.g., Churchland and Sejnowski 1990, p. 229), and Figure 6.2.1 suggest, then it is perfectly reasonable to expect elimination in those interlevel situations involving significant incommensurability. The problem, though, is that neither the history of science, nor current scientific practice, nor the scientific research the Churchlands champion, nor a concern for explanatory pluralism offers much reason to expect theory elimination in such settings.

Incommensurability in interlevel contexts neither requires the elimination of theories on principled grounds nor results in such eliminations in fact. Admittedly, in the early stages of a science's history it is not always easy to distinguish levels of analysis and, consequently, to distinguish what would count as an interlevel, as opposed to an intralevel, elimination. Crucially, though, the history of science and especially the history of late nineteenth and twentieth century science offer no examples of large-scale interlevel theory elimination (particularly of the wholesale variety standard eliminativism and co-evolution$_S$ envision) once the upper level science achieves sufficient historical momentum to enjoy the accoutrements of other recognized sciences (such as characteristic research techniques and instruments, journals, university departments, professional societies, and funding agencies). The reason is simple enough. Mature sciences are largely defined by their theories and, more generally, by their research traditions (Laudan 1977), hence, elimination of an upper level theory by a lower level theory may risk the elimination of the upper level scientific enterprise! (Presumably, this is why Nagel always spoke of the reduction of a *science*, rather than of a theory, when addressing interlevel cases.)

A motive for undertaking interlevel investigation (especially when the intertheoretic connections are not plentiful) is to explore one science's successful problem solving strategies as a means of inspiring research, provoking discoveries, and solving recalcitrant problems at another level.

(Bechtel and Richardson 1993 focus in particular on the problem of understanding the operation of mechanisms.) Monitoring developments in theories at neighboring levels is often a fruitful heuristic of discovery. The strategy's fruitfulness depends precisely on the two sciences maintaining a measure of independence from one another.

This is the mark of explanatory pluralism and co-evolution$_P$. A paucity of interlevel connections only enhances the (relative) integrity and autonomy of the upper level science. As Wimsatt notes "in interlevel reduction, the more difficult the translation becomes, the more *irreplaceable* the upper level theory is! It becomes the only practical way of handling the regularities it describes" (1976, p. 222). The theories at the two levels possess different conceptual and explanatory resources, which underscore different features of their common explanandum. They provide multiple explanatory perspectives that should be judged on the basis of their empirical success--not on hopes about their putative promise for the theoretical (or ontological) unification of science. For the pragmatically inclined, explanatory success is both sufficiently valuable and rare that it would be imprudent to encourage the elimination of any potentially promising avenue of research. As Churchland and Sejnowski remark, "the co-evolutionary advice regarding methodological efficiency is 'let many flowers bloom'" (1992, p. 13).

The Churchlands have argued famously, though, that folk psychology is barren (P.S. Churchland 1986, pp. 288-312 and P.M. Churchland 1989, pp. 2-11). Those arguments have provoked an entire literature in response (see Greenwood 1991 and Christensen and Turner 1993). I am sympathetic with the Churchlands' arguments, at least when they wield them against positions in the philosophy of mind that deny the explanatory goals and the conjectural and fallible character of folk psychology. That folk psychology offers explanations and that it is conjectural and fallible are both correct. That is just not the whole story, though.

The pivotal question for a pragmatist is whether folk psychology can contribute to the progress of our knowledge, or, better, whether folk psychology contains resources that may aid subsequent, more systematic psychological theorizing. Attribution theory, the theory of cognitive dissonance, and other proposals within social psychology employ as rich versions of the propositional attitudes as does folk psychology (Bechtel and Abrahamsen 1993). Moreover, as Dennett (1987) has emphasized, employing the intentional stance aids theorizing about operative subsystems in sub-personal cognitive psychology.[9] These are just two fronts where *psychological* science seems to be simultaneously employing and, ever so gradually, *transforming* familiar folk psychological notions. Arguably, then, the Churchlands may have underestimated the possible contribution of the resources of folk psychology, because they have been insufficiently attentive to their role in social psychological and cognitive theorizing (McCauley 1987, 1989). Indeed, they *sometimes* disregard the psychological altogether.[10] (See, however, note 14 below.)

I suspect that such neglect is born of insisting on a unified account of intertheoretic relations and of entertaining images of co-evolution$_S$, in particular. The Churchlands are correct to emphasize the salient role of theory elimination in scientific progress, but these eliminations are *intra*level processes and most univocally so (1) when the levels in question concern scientific pursuits as well established as neuroscience and psychology and (2) when those levels are construed as thickly, i.e., as inclusively, as the distinction between those two sciences implies. The theories and characteristic ontologies informing Stahl's account of combustion and Young's account of the propagation of light were replaced by theories (with new ontologies) that operated at the same levels of analysis and that were identified, both now *and then*, as continuations of the research traditions associated with those levels. Elimination in science is principally an intralevel process.

That is not to assert that interlevel considerations play no role. Even with levels of analysis so thickly construed, I do *not* mean to deny that scientists' decisions at levels above and below influence theoretical developments at a given level. Nor do I wish to deny that at that targeted level such developments can involve eliminations. Rather, the critical point is that these influences are reliably *mediated* by developments in the conceptual apparatus and research practices that are associated with the research tradition of the targeted level. (See Bechtel and Richardson 1993, especially chapter 8 and Bechtel, 1996.)

If it is *construed as an explanatory construct*, then, I agree with the Churchlands that much of folk psychology may well undergo substantial revision and, perhaps, even elimination eventually.[11] What I am suggesting, though, is:
(1) that those changes will occur primarily as a result of progress within social and cognitive psychology, i.e., that they will arise as the consequence of intralevel processes within the psychological level of analysis;
(2) that, in virtue of the role of intentional attributions in the theories of social and cognitive psychology, this displacement will probably be quite gradual, i.e., that, so far, the changes are proving evolutionary, not revolutionary;
(3) that theoretical developments within those sub-disciplines of psychology will mediate whatever co-evolutionary influence neuroscience has in this outcome.

Mapping the Churchlands' continuum on to the intralevel-interlevel distinction yields the arrangement in Figure 6.2.2. It readily accommodates co-evolution$_M$ and co-evolution$_P$, but co-evolution$_S$ finds no obvious home. The point is that the interaction of psychology and neuroscience, like *all* co-evolutionary situations, is a case of *interlevel* relations. In short, co-evolution$_S$ embodies a category mistake. It conflates the dynamics of the co-evolution of theories at different levels of analysis with those of scientific revolutions, which are intralevel processes.[12]

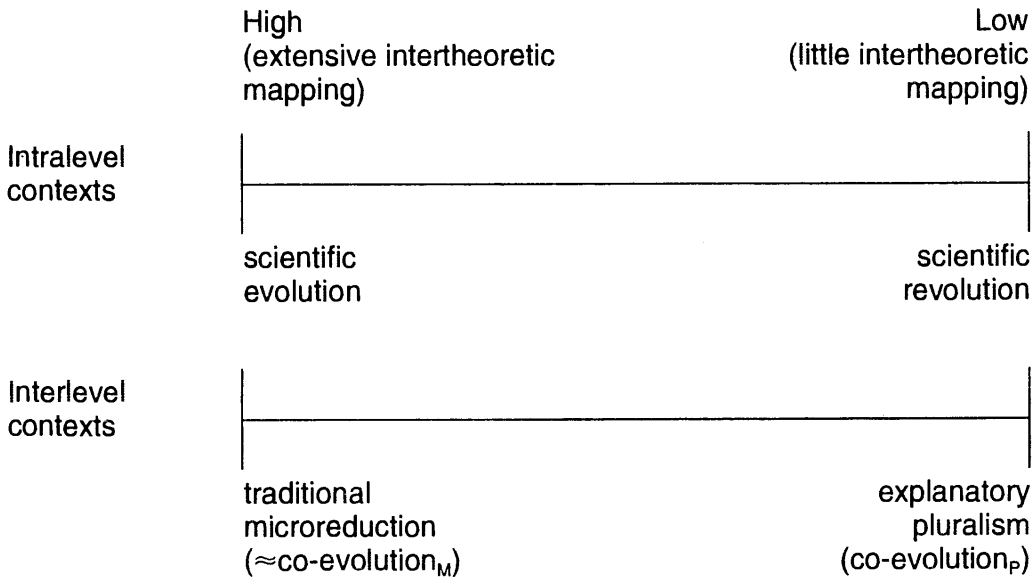


Figure 6.2.2. Mapping degrees of intertheoretic continuity (the Churchlands' continuum) onto intralevel and interlevel contexts.

What follows on this revised picture in Figure 6.2.2 about interlevel cases that reflect substantial incommensurability? In fact, I think such cases are extremely rare, especially if the sciences in

question are well established, since part of becoming a well-established science is precisely to possess theories that recognizably cohere with at least some features of theories at contiguous levels. Arguably, the distinctions between levels of analysis already presume the extreme improbability of such radical incompatibility between theories operating at adjoining *scientific* levels. (Of course, not all explanatory theories are scientific theories.) If analyses diverge in nearly all respects, then it may no longer be clear that they share a common explanandum – tempting some researchers to adopt obscurantist strategies of metaphysical extravagance.[13]

The problems surrounding co-evolution$_S$ notwithstanding, in elaborating Wimsatt's metaphor of the co-evolution of theories the Churchlands have fundamentally reinvigorated the study of change in interlevel relations over time (arguably initiated in Schaffner 1967). As with this section, the next will say more about co-evolution$_P$ by opening with more about what it is not.

## 5. Exploring Explanatory Pluralism: Beyond Co-evolution$_m$

The demand in science for coherence of theories at adjacent levels of analysis is an additional motive, beyond the promise of new discoveries, for probing possible interlevel connections. The motive is to ascertain whether or not research at nearby levels coheres with and supports scientists' findings, and if it does not, to explore possible adjustments to increase the probability of such mutual support. This can, among other things, clarify respects in which the two sciences share a common explanandum.

In the long term scientists' concern for coherence among their results inevitably tends to encourage better intertheoretic mapping in interlevel settings. Forging such connections produces new discoveries in the respective sciences. One strategy, though certainly not the only one, is to advance hypothetical identities between theoretical ontologies in order to power an engine of discovery. The relationship between Mendelian genetics and biochemical genetics over the first half of this century is an especially apt illustration of two related research programs at neighboring levels of analysis aiding one another through the investigation of a series of proposals about which structures were, in fact, the genes. Scientists' two primary motives for inquiries into research at neighboring levels, then, are finally one and the same. This might seem to suggest that co-evolution$_M$ predominates; however, a number of countervailing considerations (some of which are briefly examined in this section) favor an explanatory pluralism where the sciences maintain some independence of theory, method, and practice. So, even approximate microreduction need not be inevitable.

Two issues especially distinguish co-evolution$_M$ and co-evolution$_P$. The first concerns the relative metaphysical, epistemic, and/or explanatory priority of upper and lower level theories in the co-evolutionary process. The second concerns the grounds offered for any disparate assignments of these priorities.

The default assumption adopted in an analysis of co-evolution$_M$ that accords with the traditional microreductionistic rationale for physicalism attributes comprehensive priority to lower levels. Classical microreduction would forecast a co-evolutionary process where the overwhelming majority of the selection pressures are exerted from the bottom up. The upper level theory may contribute in the process of discovery, providing an initial vocabulary and problems for research, but sooner or later it must conform to the lower level theory's expectations. Here the grounds for this priority rest not merely on the theoretical maturity and superior precision lower level theories typically enjoy (with which pragmatism has no complaint) but also on presumptions about those theories' metaphysical preeminence. (See note 10 above.)

Occasionally[14], the Churchlands seem to subscribe to a version of co-evolution$_M$ that resembles this position. For example, Churchland and Sejnowski emphasize "the importance of the single neuron models [among the various sub-levels of analysis within neuroscience] as *the bedrock and fundament* into which network models *must* eventually fit" (1992, p. 13, emphasis added).[15] Although the Churchlands have avoided the traditional microreductionists' fervor about the replaceability of the reduced theory at the upper level (e.g., Churchland and Churchland 1990, p. 256), their repeated emphasis on lower level theories' corrections of upper level theories also suggests that selection pressures are largely unidirectional, especially when they treat these lower level elaborations as of a piece with corrections in intralevel contexts where substantial ontological modification *is* sometimes part of the package.

Co-evolution$_M$ will prove relevant to but a small percentage of cases, at best. On the one hand, if co-evolution$_M$ is supposed to issue in the classical microreductionist program (presumably, it is not), then all of the familiar objections and caveats apply–plus at least one important additional one. The sort of tight integration with a dominant lower level theory to which classical microreduction aspires must inevitably restrict research at the higher level. If there ever was a microreduction that conformed to all of the logical and ontological constraints imposed by the classical model, for example Causey's (1977) version, it would endow the lower level with an explanatory and metaphysical priority that would discourage all motives for theoretical novelty at the higher level. It would encourage only those paths of research at the higher level that promised to preserve its tight fit with the theory at the lower level. Its effect, in short, would be to check imaginative scientific proposals.

On the other hand, if co-evolution$_M$ is supposed to result only in the weaker analogical relation to which the Churchlands' model of approximate reduction looks, then the points of reductive contact may prove less extensive than the knitting metaphor suggests, and the microreductionist case for the explanatory, epistemic, and metaphysical priority of lower levels ends up seeming somewhat less compelling, especially once we have teased apart the differences in the "corrections" that occur in interlevel and intralevel contexts.

The case for co-evolution$_P$, however, does not turn exclusively on the problems the two competing conceptions face. Scrutiny of actual cases, including those in cognitive neuroscience to which the Churchlands have devoted particular attention, strongly suggests that the outcome of the co-evolution of theories is usually as co-evolution$_P$ describes. Instead of driving inexorably toward comprehensive theoretical and practical integration where the lower level theory governs, scientific opportunism is usually closer to the truth in most interlevel forays. At least initially, scientists periodically monitor developments at nearby levels searching for either interlevel support, tantalizing findings, or both.

Churchland and Sejnowski's survey of proposals concerning the neural basis of working memory is a fitting illustration (1992, pp. 297-305). Not only did the concept of "working memory" emerge out of theoretical developments in experimental psychology, but so did many of the findings that guide neural modeling. For example, Churchland and Sejnowski point explicitly to the discovery of a short term memory deficit for verbal materials in some subjects. They also highlight the ability of various interference effects both to dissociate working memory from long term memory in normal subjects and to dissociate subsystems of working memory (linked with auditory, visuospatial, and verbal materials) from one another. These discoveries in experimental psychology provided both inspiration and direction for neural modeling. They also constitute a set of findings that any relevant neuroscientific proposal should make sense of.

On even the most exacting philosophical standards, this last consideration is *epistemically* significant. Theoretical proposals and the research they spawn at the higher level do not merely contribute to the process of discovery at the lower level. The upper level science provides a body of *evidence* against which the science at the lower level can evaluate competing models. This evidence is particularly useful, precisely because it frequently arises *independently* of the formulation of the specific lower level models to whose assessment it contributes. It helps to assure the independent testability of the models in question.

It has been widely conceded that upper level theories can play a catalytic role in the process of *discovery* at the lower level. Indeed, sometimes the conceptual resources and research techniques of a lower level science are basically insufficient to enable practitioners even to recognize some of that level's fundamental phenomena without aid and direction from an upper level science. (Lykken et al. (1992) constitutes a particularly intriguing, recent illustration.) In the previous section we also saw how microreductionistic proposals to subordinate upper level explanations to lower level explanations risk needlessly downplaying valuable resources for dealing with the often huge computational burdens lower level theories entail. Upper level theorizing (e.g., in transmission genetics) contributes usefully to everyday scientific *problem solving*, even after lower level research (e.g. in molecular genetics) indicates the microlevel story is far more complicated. Scientific endeavors at different levels regularly display what Robert Burton (1993) has called a "strategic interdependence." Now we can see that upper level theorizing also initiates research that can contribute to lower level developments pertaining directly to *justification*. Microreductionistic proposals to subordinate upper level sciences to lower level sciences either epistemically or metaphysically risk needless evidentiary impoverishment.

The value of this evidence turns precisely on the fact that the research arose within a context of scientific theorizing and investigation sufficiently removed and sufficiently autonomous of the lower level research to insure an honest check. These psychological findings do not occur in isolation. They arise in the course of on-going theorizing and research at the psychological level. Their value to neuroscience rests in part on the fact that they emanate from a tradition of psychological theorizing and experimentation that neuroscience has not dominated. This is why it is worthwhile for each level of analysis to maintain a measure of independence.

As Churchland and Sejnowski note, experimental psychology has a century of findings (and theorizing) from which neuroscientists and neurocomputational modelers may draw (1992, p. 27; see too p. 240). Nothing more clearly illustrates the sort of scientific opportunism explanatory pluralism envisions than one of Sejnowski and Rosenberg's papers (1988) in defense of the claim that NETtalk plausibly models operative processes in human learning and cognition. (It is a fair question at what level of analysis connectionist modeling should be located. On the criteria I identified in section 2, it seems to occur at a level below that of social and cognitive psychology. Churchland and Sejnowski clearly regard it as a form of neurocomputational modeling. It is worth noting that Michael Gazzaniga places Sejnowski and Rosenberg's (1988) in the first half of his book, which concerns "*Neurobiologic* Considerations in Memory Function" rather than in the second half, which concerns "Psychological Dimensions of Memory Function in Humans.")
NETtalk is a connectionist system that converts English text into strings of phonemes. (Sejnowski and Rosenberg 1987) It is a three layer, feed-forward network that employs the standard back-propagation learning algorithm. On any given trial NETtalk receives seven inputs corresponding to a window of seven letters (including punctuation or spaces between words, if they happen to arise). The desired output is the correct phoneme associated with the fourth item in the window. The

three places on either side of the fourth item provide the network with information about how context affects pronunciation.

NETtalk's performance is nothing short of remarkable. It captures most of the regularities in English pronunciation and many of the irregularities as well. After 50,000 training trials with words, its accuracy with phonemes approaches 95 percent and it is virtually perfect with stresses and syllable boundaries.

The critical question for now, though, is what evidence Sejnowski and Rosenberg might cite to support the claim that NETtalk models processes that resemble those involved in human learning and cognition. A model of co-evolution as explanatory pluralism suggests that attention to the findings of experimental psychology might prove just as helpful here as attention to research on neural structure, and, in fact, not only do Sejnowski and Rosenberg look to psychology, they look to one of those century old findings about *remembering*, viz., the spacing effect.

The spacing effect is the finding that distributed practice with items enhances the probability of their long term retention more than massed practice does. If occasions for rehearsal are spaced out over time, the probability is high that memory performance will exceed that from employing some small number of massed practice sessions of comparable duration at the outset. Massed repetition facilitates memory when retention intervals are extremely short. In practical terms, the spacing effect is why cramming for an exam is not nearly so helpful as regular, daily preparation, whereas retention of two new telephone numbers supplied by Directory Assistance requires immediate, massed rehearsal, if they cannot be written down.

In the course of investigating the various hypotheses psychologists have offered for explaining the spacing effect, researchers have demonstrated its robustness across a huge variety of experimental settings, materials, and tasks. Thus, Sejnowski and Rosenberg suspect that it reflects "something of central importance in memory" (1988, p. 163). Consequently, it is by no means trivial, if NETtalk can be induced to exhibit the spacing effect. It would be even more striking, if its exhibition of the effect was similar in form to documented human performance.

Because of NETtalk's architecture the obvious comparison is with studies of cued recall. Sejnowski and Rosenberg chose a design after Glenberg (1976). The design called for training NETtalk up in the standard fashion, and then presenting it with the cues from a list of twenty paired associates where those cues were strings of six random letters and their associated responses were randomly generated phoneme and stress strings six characters long. (This insured that NETtalk's performance at the test could not be a function of any information it had acquired about English pronunciation.) During both the spacing interval between training opportunities and the retention interval before the test, NETtalk was presented with English distractor words that were part of its original training corpus. Both training on the paired associates and distractor episodes included feedback via back propagation. The order of the presentations to NETtalk in the experiment was as follows:
(1) 2, 10, or 20 presentations of each of the twenty paired associate cues;
(2) a spacing interval of 0, 1, 4, 8, 20, or 40 distractors;
(3) 2, 10, or 20 re-presentations of each of the twenty paired associate cues;
(4) a retention interval of 2, 8, 32, or 64 distractors;
(5) a test of NETtalk's accuracy in cued recall of the twenty paired associates.
In short, NETtalk displayed the spacing effect: "A significant spacing effect was observed in NETtalk: Retention of nonwords after a 64-item retention interval was significantly better when presented at the longer spacings (distributed presentation) than at the shorter spacings. In addition, a

significant advantage for massed presentations was found for short-term retention of the items" (Sejnowski and Rosenberg 1988, p. 167). Moreover, although direct comparison was impossible, NETtalk's overall response profile resembled that of Glenberg's human subjects.

The interlevel interaction here benefits both cognitive psychology and neurocomputational modeling. Sejnowski and Rosenberg briefly review the two major theoretical proposals for explaining the spacing effect in cognitive psychology, pointing out that neither the encoding variability hypothesis (e.g., Bower 1972) nor the processing effort hypothesis (e.g., Jacoby 1978) can account for all of the available data. They then suggest a further hypothesis focusing on the form in which information is encoded in a connectionist network, i.e., on the form of the memory representation. They propose that the short-term advantage of massed practice and, particularly, the longer term advantage of distributed practice are at least partially explicable in terms of the dynamics of connectionist nets.

Crucially, Sejnowski and Rosenberg do *not* construe their hypothesis as competing with (let alone correcting or eliminating) the two psychological proposals. (They have, after all, explored but one set of findings concerning cued recall.) Instead, they emphasize its compatibility with each. They claim correctly that it offers "a different type of explanation" at "a different level of explanation" (1988, p. 170). They explicitly discuss ways in which the notions of "encoding variability" and "processing effort" could map on to the dynamics of connectionist networks. These finer grained accounts of these processes in terms of a network's operations suggest bases for *elaborating* the two hypotheses.

If the co-evolution of research in interlevel contexts yields the explanatory pluralism for which I have been plumping, then it is not only the lower level that offers the aid and comfort, nor is it only the higher level that receives it. As the neural modeling of working memory illustrates, here too psychological findings provide both evidentiary support and strategic guidance to lower level modeling of brain functioning. Sejnowski and Rosenberg remark that "those aspects of the network's performance that are similar to human performance are good candidates for general properties of network models" (1988, p. 171). Their project reflects a general strategy for the testing and refinement of neurocomputational models that relies on the relative independence of work in experimental psychology. Features of particular networks that enable them to mimic aspects of the human performance that psychology documents themselves deserve mimicry in subsequent modeling of human cognition.

What is especially clear about the contribution of higher levels in this example is Sejnowski and Rosenberg's explicit acknowledgement of just how far "guidance" can go. "When NETtalk deviates from human performance, there is good reason to believe that a more detailed account of brain circuitry may be necessary" (1988, pp. 172). Their comment accords nicely with the account of explanatory pluralism I have been developing. Unlike the picture of co-evolution inspired by the tradition of microreductionism, a pragmatically inspired explanatory pluralism permits no *a priori* presumptions about lower level priority. Sejnowski and Rosenberg readily allow that our psychological knowledge enjoys sufficient integrity to forcefully urge further *elaboration* of analyses of brain systems formulated at lower levels.[16] This would be no less (nor no more) a correction of the lower level theory (or its ontology) than are the lower level "corrections" of upper level theories (and their ontologies) the Churchlands have sometimes been wont to stress.

Such divergences, then, are not grounds for dismissal. They are, rather, opportunities for advance. The co-evolution of sciences (not just theories) at contiguous levels of analysis preserves the

plurality of explanatory perspectives that the distinctions between levels imply, because leaving these research traditions to their own devices is an effective means of insuring scientific progress.

**References**

Bechtel, W. (ed.) 1986: The Nature of Scientific Integration. *Integrating Scientific Disciplines*, The Hague: Martinus Nijhoff.

Bechtel, W. 1996: What should a connectionist philosophy of science look like? In R. N. McCauley (ed.), *The Churchlands and their critics*, Oxford: Basil Blackwell.

Bechtel, W. and Abrahamsen, A. A. 1993: Connectionism and the Future of Folk Psychology. In R. G. Burton (ed.), *Natural and Artificial Minds*, Albany: SUNY Press.

Bechtel, W. and Richardson, R. C. 1993: *Discovering Complexity*. Princeton: Princeton University Press.

Bickle, J. 1992: Mental Anomaly and the New Mind-Brain Reductionism. *Philosophy of Science*, 59, 217-230.

Bower. G. H. 1972: Stimulus-Sampling Theory of Encoding Variability. In A. W. Melton and E. Martin (eds), *Coding Processes in Human Memory*, Washington: V.H. Winston & Sons.

Burton, R. G. 1993: Reduction, Elimination, and Strategic Interdependence. In R. G. Burton (ed.), *Natural and Artificial Minds*, Albany: SUNY Press.

Causey, R. 1972: Uniform Microreductions. *Synthese*, 25, 176-218.

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

Christensen S. M. and Turner D. R. (eds.) 1993: *Folk Psychology and the Philosophy of Mind*. Hillsdale, New Jersey: Lawrence Erlbaum Associates.

Churchland, P. M. 1979: *Scientific Realism and the Plasticity of Mind*. Cambridge: Cambridge University Press.

Churchland, P. M. 1989: *A Neurocomputational Perspective: The Nature of Mind and the Structure of Science*. Cambridge: The MIT Press.

Churchland, P. M. and Churchland, P. S. 1990: Intertheoretic Reduction: A Neuroscientist's Field Guide. *Seminars in the Neurosciences*, 2, 249-256.

Churchland, P. S. 1986: *Neurophilosophy*. Cambridge: The MIT Press.

Churchland, P. S., Koch, C., and Sejnowski, T. J. 1990: What is Computational Neuroscience? In E. L. Schwartz (ed.), *Computational Neuroscience*, Cambridge: The MIT Press.

Churchland, P. S. and Sejnowski, T. J. 1990: Neural Representation and Neural Computation. In W. Lycan (ed.), *Mind and Cognition: A Reader*, Oxford: Basil Blackwell.

Churchland, P. S. and Sejnowski, T. J. 1992: *The Computational Brain*. Cambridge: The MIT Press.

Dennett, D. C. 1987: Three Kinds of Intentional Psychology. *The Intentional Stance,* Cambridge: The MIT Press.

Enc, B. 1983: In Defense of the Identity Theory. *Journal of Philosophy*, 80, 279-298.

Feyerabend, P. K. 1962: Explanation, Reduction, and Empiricism. In H. Feigl and G. Maxwell (eds), *Minnesota Studies in the Philosophy of Science, Volume III*, Minneapolis: University of Minnesota Press.

Fodor, J. A. 1975: *The Language of Thought*. New York: Thomas Y. Crowell Company.

Glenberg, A. M. 1976: Monotonic and Nonmonotonic Lag Effects in Paired-Associate and Recognition Memory Paradigms. *Journal of Verbal Learning and Verbal Behavior*, 15, 1-16.

Greenwood, J. D. (ed.) 1991: *The Future of Folk Psychology*. New York: Cambridge University Press.

Hirst, W. and Gazzaniga, M. 1988: Present and Future of Memory Research and Its Applications. In M. Gazzaniga (ed.), *Perspectives in Memory Research*, Cambridge: The MIT Press.

Hooker, C. 1981: "Towards a General Theory of Reduction," Dialogue 20: 38-59, 201-36, 496-529.

Jacoby, L. L. 1978: On Interpreting the Effects of Repetition: Solving a Problem Versus Remembering a Solution. *Journal of Verbal Learning and Verbal Behavior*, 17, 649-667.

Kuhn, T. 1970: *The Structure of Scientific Revolutions* (2nd edition). Chicago: University of Chicago Press.

Laudan, L. 1977: *Progress and Its Problems*. Berkeley: University of California Press.

Lehky, S. R. and Sejnowski, T. J. 1988: Network Model of Shape-from-Shading: Neural Function Arises from Both Receptive and Projective Fields. *Nature*, 333, 452-454.

Lykken, D. T., McGue, M., Tellegen, A., and Bouchard, T. J. 1992: Emergenesis: Genetic Traits That May Not Run in Families. *American Psychologist*, 47, 1565-1577.

McCauley, R. N. 1981: Hypothetical Identities and Ontological Economizing: Comments on Causey's Program for the Unity of Science. *Philosophy of Science*, 48, 218-227.

McCauley, R. N. 1986: Intertheoretic Relations and the Future of Psychology. *Philosophy of Science*, 53, 179-199.

McCauley, R. N. 1987: The Role of Cognitive Explanations in Psychology. *Behaviorism* (subsequently *Behavior and Philosophy*), 15, 27-40.

McCauley, R. N. 1989: Psychology in Mid-Stream. *Behaviorism* (subsequently *Behavior and Philosophy*), 17, 75-77.

McCauley, R. N. 1993: Brainwork: A Review of Paul Churchland's *A Neurocomputational Perspective*. *Philosophical Psychology*, 6, 81-96.

McCauley, R. N. (forthcoming): Cross-Scientific Relations: Toward an Integrated Approach to the Study of the Emotions. In B. Shore and C. Worthman (eds), *The Emotions: Culture, Psychology, Biology*.

Nagel, E. 1961: *The Structure of Science*. New York: Harcourt, Brace and World.

Neisser, U. 1967: *Cognitive Psychology*. New York: Appleton-Century-Crofts.

Oppenheim, P. and Putnam, H. 1958: Unity of Science as a Working Hypothesis. In H. Feigl, M. Scriven, and G. Maxwell (eds), *Minnesota Studies in the Philosophy of Science--Volume II*, Minneapolis: University of Minnesota Press.

Richardson, R. 1979: Functionalism and Reductionism. *Philosophy of Science*, 46, 533-558.

Schaffner, K. 1967: Approaches to Reduction. *Philosophy of Science*, 34, 137-147.

Sejnowski, T. J. and Churchland, P. S. 1989: Brain and Cognition. In M. Posner (ed.), *Foundations of Cognitive Science*, Cambridge: The MIT Press.

Sejnowski, T. J. and Rosenberg, C. R. 1987: Parallel Networks that Learn to Pronounce English Text. *Complex Systems*, 1, 145-168.

Sejnowski, T. J. and Rosenberg, C. 1988: Learning and Representation in Connectionist Models. In M. Gazzaniga (ed.), *Perspectives in Memory Research*, Cambridge: The MIT Press.

Thagard, P. 1992: *Conceptual Revolutions*. Princeton: Princeton University Press.

Wimsatt, W. C. 1976: Reductionism, Levels of Organization, and the Mind-Body Problem. In G. Globus, G. Maxwell, and I. Savodnik (eds), *Consciousness and the Brain*, New York: Plenum Press.

---

1. The story is even more complex, since each level of analysis has both a synchronic and a diachronic moment for which separate theories have been developed. See McCauley (forthcoming). At the biological level, for example, cell biology is one of the synchronic sub-disciplines focusing on the structures within the cell whereas evolutionary biology is devoted to the study of change in forms of life over time. The Churchlands' have confined their discussions almost exclusively to synchronic examples.

2. One of the first, if not *the* first, is Wimsatt's (1976) classic discussion.

3. Although they concur with Churchland's judgment that the folk psychological notion of a unitary faculty of memory is probably wrong, Hirst and Gazzaniga (1988, pp. 276, 294, and 304-05) seem to adopt a far more sanguine view about the contributions of psychology (both folk and experimental) to our understanding of memory. They recognize that the fragmentation of 'memory' need not lead to its elimination. (See section 5 below.)

4. . . . and the position from which they have generally (though not unequivocally) retreated over the past few years.

5. See Churchland and Sejnowski 1990, p. 229, Churchland, Koch, and Sejnowski 1990, pp. 51 and 54, and Churchland and Sejnowski 1992, pp. 10-13.

6. Consider the discussion in Neisser (1967).

7. Interestingly, Ernest Nagel's *The Structure of Science* (1961), the *locus classicus* of traditional research on reduction, implicitly recognizes the importance of distinguishing between intralevel and interlevel contexts. Nagel consistently describes intralevel cases as involving the reduction of *theories* and interlevel cases as involving the reduction of *sciences*.

8. This is, *in part*, the result of the same considerations that motivate the Churchlands and Richardson's (1979) arguments that alleged reductions that conform to traditional microreductionistic standards can only be domain specific.

9. –or in neuropsychology, as Churchland and Sejnowski's (1992, p. 282) discussion of the role of the hippocampus in short term memory illustrates. See P.S. Churchland 1986, p. 361.

10. An interesting illustration arises in Churchland and Sejnowski's discussion of the major levels of organization in the nervous system (1992, pp. 10-11). Their diagram of the relevant levels tops out at the central nervous system with no mention of psychology. The obvious defense is to note that the diagram addresses *anatomical* structures of the nervous system only. Fair enough. What is telling, though, is a footnote (1992, p. 11, footnote 5) to this discussion. Churchland and Sejnowski concede that a more comprehensive account would include a *social* level above the central nervous system. At least for the purposes of this discussion, they seem not even to countenance the possibility that cognitive research may capture organizational structure of explanatory significance not immediately reducible to the neurophysiological. (See too Sejnowski and Churchland 1989, p. 343.)

A meta-level comment: the physicalist holds that metaphysical manifestness (which, remember, is *physical* manifestness for the physicalist) constrains what will count as *satisfactory* explanation, whereas the pragmatist proposes that explanatory success *should* constrain metaphysical commitment. If that diagnosis is correct, the on-going negotiation in the Churchlands' work I described in section 3 is, at its root, one about competing norms.

11. I should emphasize that I am speaking of the elimination of folk psychology as an explanatory construct *within* scientific psychology. The elimination of the principles of folk physics centuries ago in physics has had little effect on its persistence among the folk.

12. The illustrations the Churchlands (1990) offer in support of their “overview of the general nature of intertheoretic reduction” (p. 249) proceed in the following order:

(1) the reduction of Kepler's laws to Newton's (intralevel);

(2) the reduction of the ideal gas law to the kinetic theory--emphasizing (p. 250, some emphasis added) that "this reduction involved *identifying* a familiar *phenomenal* property of common objects with a highly unfamiliar *micro*-physical property" (interlevel);

(3) the reduction of classical (valence) chemistry by atomic and sub-atomic (quantum) physics (interlevel);

(4) the reduction of Newtonian mechanics to the mechanics of Special Relativity (intralevel);

(5) the elimination of phlogiston by Lavoisier's oxygen theory of combustion (intralevel).

13. But just as progress in tracing the relevant biological systems preserved the vitality of organisms without vitalism, so too is progress at tracing the relevant psychological systems slowly revealing how we can preserve the cleverness and wondrous experiences of intelligent creatures without dualism. The interlevel influences of neuroscience will no more co-opt or eliminate psychological theorizing than the interlevel influences of chemistry co-opted or eliminated physiological theorizing.

14. As noted near the end of section III, the Churchlands more often seem to endorse an account of co-evolution resembling co-evolution$_P$. In Churchland and Sejnowski 1990 (p. 250) and 1992 (p. 240), they not only advocate a form of explanatory pluralism, but they explicitly include the psychological sciences.

15. Conceding that it will not involve a single model nor direct explanations of higher levels in terms of events at the molecular level, Churchland and Sejnowski, nonetheless, aspire to a "unified account" of the nervous system, where "the integration [will] consist of a chain of theories and models that links adjacent levels" (Sejnowski and Churchland 1989, p. 343).

16. If neurocomputational modeling of networks constitutes a higher level of analysis than does the study of particular neurons (and it certainly seems to on Churchland and Sejnowski's view – 1992, p. 11), then Churchland and Sejnowski's (1992, pp. 183-188) take on recordings of single cells' response profiles in the visual cortex is an illustration of just the sort of circumstances that the Sejnowski and Rosenberg citation allows for--one in which higher level research impels a reevaluation of lower level doctrines.

Churchland and Sejnowski (following Lehky and Sejnowski 1988) argue that neurocomputational research on the visual system's ability to extract shapes exclusively from information about shading reveals that the conventional interpretation of the function of receptive fields of neurons in the visual cortex may well be wrong. That interpretation, which arose from single cell studies, holds that these neurons function as edge and bar detectors. Churchland and Sejnowski maintain that this interpretation ignores the cells' projective fields. Hidden units in Lehky and Sejnowski's model developed receptive fields with similar response profiles, however these orientations were the result of training the network on the shape from shading task. "In a trained-up network, the hidden units represent an intermediate transformation for a computational task quite different from the one that has been customarily ascribed . . . they

are used to determine the shape from the shading, not to detect boundaries" (Churchland and Sejnowski 1992, pp. 185-186).