[Millikan, 1993] R. G. Millikan. *White Queen Psychology and Other Essays for Alice.* Cambridge, MA: MIT Press, 1993.
[Neander, 1991] K. Neander. Functions as Selected Effects, *Philosophy of Science* 58: 168-184, 1991.
[Place, 1956] U. T. Place. Is Consciousness a Brain Process?, *The British Journal of Psychology* 47: 44-50, 1956.
[Polger, 2004] T. Polger. Neural Machinery and Realization, *Philosophy of Science* 71: 997-1006, 2004.
[Putnam, 1960] H. Putnam. Minds and Machines, 1960. Reprinted in his *Mind, Language, and Reality: Philosophical Papers, Volume 2.* New York: Cambridge University Press, 1975.
[Shapiro, 2000] L. Shapiro. Multiple Realizations, *Journal of Philosophy* 97: 635-654, 2000.
[Shapiro, 2004] L. Shapiro. *The Mind Incarnate.* Cambridge, MA: MIT Press, 2004.
[Smart, 1959] J. J. C. Smart. Sensations and Brain Processes, 1959. Reprinted in D. Rosenthal (editor), *The Nature of Mind.* New York, Oxford University Press, 1991.
[Sober, 1985] E. Sober. Putting the Function Back into Functionalism, 1985. Reprinted in W.G. Lycan (editor), *Mind and Cognition: An Anthology.* Oxford: Blackwell, 1998, 2nd edition.
[von Melchner *et al.*, 2000] L. von Melchner, S. Pallas, and M. Sur. Visual Behaviour Mediated by Auditory Cortex Directed to the Auditory Pathway, *Nature* 404: 871-876, 2000.
[Wilson, 2001] R. A. Wilson. Two Views of Realization, *Philosophical Studies* 104: 1-31, 2001.
[Wilson, 2004] R. A. Wilson. *Boundaries of the Mind: The Individual in the Fragile Sciences: Cognition.* New York: Cambridge University Press, 2004.
[Wilson and Keil, 1998] R. A. Wilson and F. C. Keil. The Shadows and Shallows of Explanation, 1998. Modified version reprinted in F.C. Keil and R.A. Wilson (editors), *Explanation and Cognition.* Cambridge, MA: MIT Press, 2000.
[Wilson and Keil, 1999] R. A. Wilson and F. C. Keil, eds. *The MIT Encyclopedia of the Cognitive Sciences.* Cambridge, MA: MIT Press, 1999.
[Wimsatt, 1997] W. Wimsatt. Aggregativity: Reductive Heuristics for Finding Emergence *Philosophy of Science* 64: S372-S384, 1997.

REDUCTION: MODELS OF CROSS-SCIENTIFIC RELATIONS AND THEIR IMPLICATIONS FOR THE PSYCHOLOGY-NEUROSCIENCE INTERFACE

Robert N. McCauley

1 INTRODUCTION

With the rise of functionalism and of new proposals about putative intrinsic properties of the physical necessary to account for conscious experience, enthusiasm among philosophers about the possibilities of reducing the theories of psychology to those of neuroscience probably reached its nadir in the 1990s. Ned Block [1997] noted that recent philosophy of psychology has witnessed (very nearly) an "antireductionist consensus," and Jaegwon Kim suggested that this consensus extended well beyond the confines of professional philosophy:

> ... "reduction," "reductionism," "reductionist theory," and "reductionist explanation" have become pejoratives not only in philosophy, on both sides of the Atlantic, but also in the general intellectual culture of today. They have become common epithets thrown at one's critical targets to tarnish them with intellectual naivete and backwardness. ... If you want to be politically correct in philosophical matters, you would not dare come anywhere near reductionism, nor a reductionist [1998, p. 89].

The widespread presumption is that proposed reductions of the psychological to the neural are so obviously hopeless that when they address the deepest philosophical problems of mind philosophers need not trouble themselves much with the details of either scientific investigations pertaining to their connections or philosophical accounts of those investigations.

On three prominent fronts these developments seem (at the very least) unexpected from the standpoint of the philosophy of science. First and, perhaps, most important, the history of science provides no grounds for such pessimism (let alone, such a dismissive view) about the prospects for successful reductions in these environs. The explanatory triumphs of the resulting theoretical integrations have richly rewarded the eagerness with which scientists have pursued reductive

Handbook of the Philosophy of Science. Philosophy of Psychology and Cognitive Science
Volume editor: Paul Thagard
General editors: Dov M. Gabbay, Paul Thagard and John Woods

projects over the past one hundred fifty years. Reduction has probably been the single most effective research strategy in the history of modern science, engendering more precise accounts of the mechanisms (and their operations) underlying everything from magnetic forces to organisms' inheritance of traits to the visual perception of moving objects — to note but three examples from three different levels of analysis in science and three different collections of decades in the two centuries in question. Exploring reductive possibilities opens new avenues for sharing methodological, theoretical, and evidential resources. Successful reductions reliably generate productive programs of research at the analytical levels from which the candidate theories hail, squaring the lower level, mechanical details with the upper level phenomenal patterns and refining our understanding of both in the bargain.

Second, the opportunities for such theoretical integrations between the cognitive and the neural sciences have, if anything, only increased during the time period when this anti-reductionist consensus has prevailed among philosophers. The emergence in the 1960s and 1970s of the multi-disciplinary project that has become cognitive science provided example after example of researchers attempting to mine evidence from work in related disciplines and to develop cross-scientific resonances of theory. The growing use of brain-imaging technologies, especially PET and fMRI, over the next two decades has only accelerated the pace at which the multi-level theoretical proposals that have marked recent cognitive neuroscience have appeared. These proposals regularly conjoin insights from psychology, network modeling, clinical neurology, cellular neuroscience, and more. The past forty years have witnessed not only a rapid multiplication in the sheer number of findings about the mind-brain available for reductive analysis but dozens of interdisciplinary projects in the relevant sciences that have made reductive headway.

Finally, philosophers of science have diligently sought at least since the 1950s to provide general models of reductive accomplishments in science ([Kemeny and Oppenheim, 1956]; [Oppenheim and Putnam, 1958]). In the final decade of logical empiricism's reign, Ernest Nagel's [1961/1979] account of reductive relations in science (referred to hereafter as the "standard model") emerged as the touchstone for subsequent discussions among philosophers of mind. There Nagel speaks about both the reduction of scientific theories and the reduction of whole sciences. He construes both in terms of the logical derivation of a reduced theory's laws from the laws of a reducing theory, supplemented by bridge principles that specify systematic connections between the two theories' predicates and the boundary conditions within which those connections hold. The anti-reductionist consensus among both functionalists and the friends of consciousness has mostly turned on arguments that the considerations they raise establish (different) barriers to reduction — as characterized by Nagel [Bickle, 1998, p. 5].

It probably comes as no surprise, though, that philosophers of science, including those interested in the psychological and neural sciences, have not ceased to advance new models of scientific reduction and of cross-scientific relations, more generally. These models differ variously and, sometimes, considerably from Nagel's

account. Virtually all of them include provisions that disarm the principal arguments on which the anti-reductionist consensus rests. Although they differ both in detail and in the strength of the reductionism they defend, they broadly concur that reports of the death of positions that explain mentality on the basis of neural operations and that identify features of minds and brains have been greatly exaggerated and that the character of conscious experience does not constitute an insuperable obstacle to proposing such hypothetical identities.

Section 2 briefly sketches Nagel's standard model of reduction and then discusses how the machinery of the "New Wave" model of reduction has transformed one of the standard model's principal problems into a virtue. Section 3 explicates the New Wave continuum of comparative goodness of intertheoretic mapping. Section 4 situates within the New Wave framework the two major arguments informing the anti-reductionist consensus among recent philosophers of mind. These arguments concern the multiple realizability of mental states and the irreducibility of conscious experience. Sections 5 and 6 review criticisms of the New Wave model suggesting that its proximity to the logical empiricist model on two fronts renders it, first, insufficiently sensitive to the wide range of cross-scientific relations that arise and, second, capable of engendering misleading conclusions about the status and fate of the cognitive and psychological sciences relative to neuroscience. Section 7 presents a more fine-grained model of intertheoretic relations that distinguishes between two major sorts of cases that the New Wave models lump together. Coincident with work on mechanistic explanation in science (discussed in section 9), this alternative analysis contrasts two sorts of cases that exhibit diverging profiles and considers New Wave counter-arguments against distinguishing them. It elucidates the explanatory pluralism that dominates in cross-scientific settings. Section 8 suggests that, although disagreements about general models of scientific reduction persist, confluences of opinion have emerged over the last few years in the works of philosophers interested in exploring fruitful cross-scientific relations at the borders between the cognitive and neural sciences. The first concerns the distance of functionalists' multiple realizability argument from the practices and discoveries of working scientists in these fields. Section 9 takes up a second sort of confluence concerning the crucial role that mechanistic analyses play in those practices and discoveries. Recent mechanistic analyses apply the morals of explanatory pluralism to models for the detailed study of particular patterns and mechanisms in nature. These positions rule out the most ambitious aims of New Wave reductionists in interlevel settings. Finally, section 10 examines how defusing the multiple realizability argument and taking a closer look at the practices and discoveries of scientists working in these areas suggests that arguments for the unique character of conscious experience are largely irrelevant to the sorts of considerations that lead scientists to hypothesize intertheoretic identities in reductive contexts. The Heuristic Identity Theory incorporates these insights about cross-scientific research, advancing a new, more scientifically informed, version of the psycho-physical identity theory.

2 THE STANDARD MODEL OF REDUCTION AND NEW WAVE REVISIONISM

Scientific reduction, according to Nagel and the logical empiricists, is a deduction of the laws of one scientific theory (the reduced theory) from those of another (the reducing theory). This inference requires supplementing the laws of the reducing theory with a set of ancillary statements that lay out systematic connections between the two theories' predicates while incorporating the boundary conditions within which those connections are realized. (See Wilson and Craver's "Realization" in this volume.) On this account reductions are a type of explanation in which, unlike most cases of scientific explanation, the explanandum is *not* some phenomenon but rather some law or other of the theory that is being reduced. A successful reduction on this standard model demonstrates how the reducing theory's explanatory resources encompass those of the reduced theory that is to be mapped on to it. Thus, in effect, the reduced theory constitutes an application of the reducing theory in one of its sub-domains that the boundary conditions specify.

Philosophers have used a variety of phrases ("bridge principles," "reduction functions," "coordinating definitions," etc.) to refer to the ancillary statements that supplement the laws of the reducing theory, and they have offered various proposals about the connections those statements should establish between the two theories' predicates. The significant point for now is that on the standard model those connections must enable the reduction to meet two important constraints. The first constraint is logical; the second is material.

The first constraint is concerned with assuring the "derivability" of the reduced theory from the reducing theory. In order for the reducing theory to explain the reduced theory, the latter must follow from the former (supplemented by the bridge principles) as a deductive consequence. That is because explanation for the logical empiricists conforms to the deductive-nomological (D-N) model. On the D-N model explanation involves the *derivation* of statements about what is to be explained from scientific *laws*. Consequently, the reduction functions have to articulate connections between the two theories' predicates of sufficient logical strength to support the derivation.

The second constraint is the "connectability" condition. In order for the reduction to help justify a metaphysical unity as well as a theoretical unity to science, the reduction functions also have to certify substantial connections between the entities and their properties that the two theories discuss. Establishing such connections between scientific theories motivates *programs* for unifying science via "microreductions," in which the entities the reducing theory discusses constitute the components of the entities that the reduced theory endorses ([Oppenheim and Putnam, 1958]; [Causey, 1977]). Such programs not only aim to fashion a compelling case based on mereological relations for a materialist metaphysics but also envision the reduction of entire sciences. They foresee, at least in principle, the possibility of scientists eventually abandoning research at higher levels of analysis

in deference to explanatory theories at lower levels that are simultaneously more comprehensive and more detailed.

Robert Causey's [1977] proposal for a theoretical unification of the sciences is, perhaps, the most thoroughgoing. Paul M. Churchland [1979] and Patricia S. Churchland [1986] have jointly initiated one of the best known programs exploring the possibilities for such deflationary consequences, while John Bickle [1998; 2003] offers some of the more spirited contributions along these lines recently. Bickle, for example, is concerned with "the reduction of ... psychology to neuroscience," and he holds that "reduction is a proof of displacement (in principle), showing that a typically more comprehensive theory contains explanatory and predictive resources that parallel those of the reduced theory" [1998, 214 and 28].

Proposals differ about the logical and material strength of such intertheoretic connections. The major options are (1) that the various predicates of the reducing theory constitute sufficient conditions for predicates in the reduced theory, (2) that they constitute necessary and sufficient conditions, or (3) that they not only constitute necessary and sufficient conditions but that they also involve intertheoretic identities. (See [Nagel, 1961/1979, 354–355]; [Causey, 1977], respectively.) The comprehensive mapping of the predicates applicable to the entities that the reduced theory countenances on to the predicates applicable to the entities that the reducing theory countenances vindicates assumptions about correspondences between the two theories' ontologies.

All three of these options possess sufficient logical muscle to underwrite the derivation that the first constraint demands, so it is primarily the problem of ascertaining what is required for adequately linking the theories' ontologies, i.e., what is required for meeting the second constraint, that has occupied subsequent commentators. How strong of an intertheoretic "mapping" relation is required to variously achieve (a) the explanation of the reduced theory? (b) the displacement of the reduced theory? and (c) the theoretical and ontological unification of science? On all three of options (1) - (3) above, the reduction functions can be regarded, in effect, as hypotheses that call for empirical support, and, correspondingly, discussions have regularly taken up the character of that support, how advocates of any particular reductive explanation might gather it, and what it would entail concerning questions (a) - (c) above.

The appeal of the standard model's formality, clarity, and precision is uncontested. Philosophers, however, began to realize that its idealized account of intertheoretic relations came at the price of its ability to capture a large range of actual cases of intertheoretic relations that did not meet its exacting requirements [Wimsatt, 1978]. Faithful portrayals of many relationships between scientific theories often resulted in candidate reduction functions that were weaker logically than the options listed above and that supported partial mappings only. Therefore, the resulting connections frequently seemed capable of sustaining neither the derivation of the theory to be reduced nor the comprehensive mapping of its ontology on to that of the putative reducing theory. Compare, for example, Patricia Churchland's diverging assessments in the 1980s of the prospects for the reduction of

various aspects of consciousness [P. S. Churchland, 1983; 1988]. Without a doubt, the most celebrated analysis of such failures of mapping was the Churchlands' profound pessimism concerning the possible connections between our everyday folk psychology and theories in neuroscience ([P. S. Churchland, 1986]; [P. M. Churchland, 1989]).

Although this diagnosis does not tally well with the logical empiricists' conceptions of reduction and the unity of science, it is consonant with the persisting impression in many cases that the reducing theory's resources do not merely encompass those of the reduced theory. If, for no other reason, on the basis of its added precision alone, the reducing theory usually appears to *improve* upon the reduced theory's account of things. Not infrequently, it corrects it. Even the familiar case of the reduction of the classical gas laws yields corrections to their predictions at extreme temperatures and pressures. Or within cognitive neuroscience itself, David van Essen and Jack Gallant's [1994] more articulated picture of the numerous connections permitting the sharing of information in the processing streams of the primate visual systems' "what" and "where" pathways, arguably, constitutes a correction of the initial proposal of Leslie Ungerleider and Mortimer Mishkin, which construed these sub-systems' operations as basically isolated from one another ([Ungerleider and Mishkin, 1982]; [Mishkin, Ungerleider, and Macko, 1983]).

If reducing theories often *correct* reduced theories in intertheoretic reductions, then, on the standard model of reduction, the reduced theories' laws should *not* follow with deductive validity from premises about the laws of the reducing theories in conjunction with the bridge principles. In dealing with some of the most impressive reductions in the history of science, advocates of the standard model find themselves faced with the embarrassing dilemma of having to repudiate the D-N model of explanation unless they will accept reduction functions that leave enough semantic slack to render the putative derivation guilty of equivocation. After all, false reduced theories cannot be validly deduced from true reducing theories, and they cannot even appear to do so unless the argument involves an equivocation. (See [Wimsatt, 1976, 218]; [P. M. Churchland, 1989, 48].)

The next generation of philosophers interested in modeling scientific reductions came to regard our *inability* to sustain reduction functions without semantic slack, i.e., our inability to formulate defensible reduction functions capable of underwriting the derivation of the reduced theory's regularities, as a *virtue* of any putative reduction that improves upon those regularities. Instead of standing by a formally perspicuous, idealized model of intertheoretic reduction that fails to describe many cases accurately, the successors of the standard model allow for the relaxation of its requirements. For example, Kenneth Schaffner [1967] argued that strictly speaking, what can be deduced from the reducing theory is not the reduced theory itself but only an *analogue* of that theory.

This proposal inspired what Bickle [1998] has dubbed the "New Wave" model of intertheoretic reduction. (Although the accounts of scientific reduction and cross-scientific relations that Bickle, the Churchlands, and Clifford Hooker propound do not coincide in every last detail, they are sufficiently similar on the fronts that

matter here that for ease of exposition I will use Bickle's "New Wave" label.) On this "New Wave" model the reducing theory does not explain the reduced theory, so the dilemma disappears. Instead, it explains an analogue of the reduced theory constructed within the conceptual framework of the reducing theory. Thus, on the New Wave view the analogical relationship between the reduced theory and its reconstruction in terms of the reducing theory's conceptual resources enables the reducing theory both to correct the reduced theory and to explain at least something very much like it at the same time. Moreover, relying on analogy, the New Wave model of reduction, apparently[1], accomplishes all of this without needing to specify bridge principles, and, therefore, without needing to explicate either their logical or ontological status. Hooker [1981] was the first to explore this proposal at length and to provide a formal explication. He notes in the course of that exposition that the strength of the analogy can vary considerably from one case to another, resulting in a spectrum of analogical strength that ranges from retentive reduction at one end to outright theory replacement at the other.

The difficulties surrounding appeals to analogical reasoning, however, are familiar. Just how close does the analogy need to be in order to justify reductive claims and how is the "closeness" of an analogy to be measured in the first place? How well must the reduced theory map on to the reducing theory in order to establish their explanatory and ontological continuity? New Wave reductionists have offered various proposals that conform to Schaffner and Hooker's general approach. For example, Paul Churchland [1989, 49] suggests that the reducing theory should provide an "equipotent image" of the reduced theory. The *equal potency* concerns its explanatory and predictive capabilities. The equipotent analogue, formulated in terms of the reducing theory's conceptual resources, should explain and predict the phenomena that the reduced theory addresses. To constitute an *image* of the reduced theory, the analogue may not have to map it comprehensively, but it should preserve that theory's principal contours from the standpoint of the causal relations it systematizes. Similarly, Bickle, who prefers to explicate reductive relations within the framework of a non-sentential, structuralist account of theories, aims to show how the analogues "mimic the structure" of the reduced theories [1998, 65].

Although talk of either images or mimicry is unlikely to meet the derivability constraint of the standard model or to point toward conceptions of explanation that square with the D-N model, they do undergird a picture of *approximate reduction* that embraces the familiar cases and, at least, offers an initial, if not an especially precise, step toward ascertaining just how much slack is tolerable. Bickle is sensitive to the fact that the cost of the New Wave models' broader

[1]Ronald Endicott [1998] argues persuasively that New Wave reduction does *not* avoid the problem of formulating satisfactory reduction functions but only relocates it. Endicott maintains that neither Paul Churchland's [1989] nor Bickle's [1998] (different) non-sentential analyses of theories, finally, enable them to avoid the fact that scientific theories always involve at least some "public-language sentences" [1998, 71] and, thus, any account of theory reduction (including these New Wave accounts) that aims to draw ontological conclusions must face the problem of specifying "a set of intertheoretic bridge laws" [1998, 72].

applicability is their vagueness. So, he [1998] employs the formal machinery of the structuralist program to provide a means for calibrating the degrees with which the theory-analogues approximate the commitments of the reduced theories.

Bickle's structuralist account characterizes theories in terms of their models and their intended empirical applications. A theory's models are the real world and mathematical systems that possess the structures it describes. Its intended empirical applications are all of the actual systems in the world to which it applies, as specified by the relevant scientific community. Models consist of (1) base sets, whose elements are classified according to the theory's categories, (2) auxiliary sets, which are abstract spaces which the theory's explanations presume, and (3) fundamental relations and functions, i.e., operations on the elements of (1) and (2). The collection of some theory's fundamental relations and functions constitutes the structure of its models. On this specific version of the New Wave account, reduction is, in effect, the mapping of particular models and their intended applications across two theories. That mapping must satisfy a variety of conditions, but the significant point is that the itemized accounting that this model theoretic approach affords permits *a measure* of the goodness of intertheoretic mapping.

On the New Wave account, the standard model's ideal designates an end point on the continuum of the comparative levels of isomorphism between reduced theories and their analogues. The continuum orders the relative goodness-of-mapping relations possible between reduced theories and their images constructed within the frameworks of their corresponding reducing theories. None of the New Wave reductionists, though, offer any precise criteria for when the amount of slack is no longer tolerable, i.e., when the theory-analogue's approximation of the reduced theory becomes too loose to make sense of reductive talk. Bickle [1998, 100–101] readily notes this limitation. At some point on that continuum the goodness-of-mapping becomes sufficiently weak that the case for intertheoretic continuity collapses.

Ironically, both friends [Fodor, 1975] and foes [P. M. Churchland, 1989] of folk psychology agree that this is the character of its relationship with the theories of neuroscience. New Wave reductionists hold that such situations make not for theory reduction but for the "historical theory succession" that marks scientific revolutions [Bickle, 1998, 101]. The superior theory simply displaces its inferior counterpart. If their intertheoretic mappings are as tenuous as those in uncontroversial historical cases such as, say, those between Stahl's account of combustion and Lavoisier's or those between Gall's phrenological hypotheses and modern cognitive neuroscience, we are, presumably, justified in speaking of the complete *elimination* of the inferior theory. Of course, it appears that the theories that risk elimination in the case at hand are, at the very least, those of folk psychology and, presumably, those in other areas of psychology that appeal to similar notions.[2]

[2]Although section 8 below will maintain that Bickle's [2003] most recent extended discussion of these matters offers some grounds for situating his position alongside philosophical treatments that do not always foresee the elimination of psychology, it is still worth noting that he characterizes his "ruthless" reductionism as one in which all such failures to map basic entities of the reduced theory result in their being "related *in a domain eliminating way*" to the machinery of the reducing theory [2003, 98]. What, apparently, has changed is not his analysis of contexts marked by mapping failures but, rather, Bickle's assessment of how much of the psychology-neuroscience interface that analysis captures.

3 THE NEW WAVE CONTINUUM OF THE COMPARATIVE GOODNESS OF INTERTHEORETIC MAPPING

The New Wave model situates different cases of intertheoretic relations at various points on a continuum of comparative goodness of intertheoretic mapping. (See figure 1.)

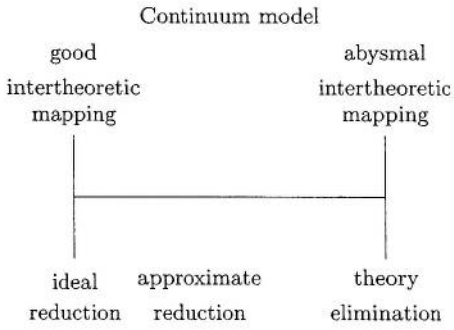

Figure 1.

These include cases situated quite near the end point of the continuum defined by the standard model's ideal, i.e., at the left end of the continuum that figure 1 portrays. Advocates of the standard model cited examples from basic physical science (such as the reduction of the wave theory of light to electromagnetic theory). Uncontroversial examples concerning the psychological and neural sciences may not exist, but, of a piece with William Wimsatt's [1978] observation, noted above, few, if any, actual cases of intertheoretic relations fully meet these exacting standards in *any* science.

New Wave theorists agree with Wimsatt's judgment. Hooker, for example, suspects that "the retention extreme of the retention/replacement continuum goes unoccupied" [1981, 45]. If even the standard model's parade cases from the physical sciences, in fact, fall at some distance on this continuum from the anchor point that designates that ideal, then that would only underscore the significance of New Wave analyses' abilities to make sense of these many familiar cases of approximate reduction. On the New Wave account the standard model's parade cases *are* only approximate reductions, since they reliably require minor counterfactual assumptions. (See [Bickle, 1998, especially p. 38 and 2003, p. 11].) Examples include the approximate reduction of the classical gas laws to principles of the kinetic theory and statistical mechanics and of Kepler's laws concerning planetary orbits

to those of Newtonian mechanics. Clearly, if this is where the most thoroughgoing reductions from the physical sciences fall on the continuum in figure 1, then the consensus of philosophical opinion would locate *all* reductive proposals linking theories from the psychological and neural sciences even further away from the standard model's ideal. Even the best of the psychology-to-neuroscience cases would be comparatively *less* approximate reductions. Candidates from these domains would include the reduction of psychological proposals about the "switch" responsible for the consolidation of declarative memories in terms of the molecular mechanisms (both subcellular and extracellular) underlying both the transition from early phase (E-LTP) to late phase (L-LTP) long term potentiation and the preservation of the latter [Bickle, 2003, ch. 2].

As bases for constructing an analogue of the reduced theory dwindle, cases are arrayed further and further to the right on the continuum in figure 1. On the New Wave account the prospects for retaining either the principles or the ontology of the theory to be reduced decrease as cases exhibit fewer and fewer correspondences. Many of the classic revolutions in the history of science fall here. These include the elimination of the Aristotelian-Ptolemaic cosmology and the impetus theory with the rise, respectively, of the Copernican system and Galileo's investigations of terrestrial mechanics. New Wave reductionists, especially the Churchlands, are famous for their arguments that many cases of intertheoretic relations at the interface of psychology and neuroscience should be located at this end of the continuum. (See [P. M. Churchland, 1989, 1–22]; [P. S. Churchland, 1986, 373] as well as [Bickle, 1998, 30, figure 2.1].)

In the first sort of case (near the left end of the continuum in figure 1) the intertheoretic mapping is delightfully smooth, and the explanatory power of the reduction is transparent. In the second sort of case (falling, say, in the left half of the continuum), the analogies are close enough and the mappings remain substantial. Any improvements or corrections at the reduced theory's edges are a function of the heightened precision the reducing theory affords. Increasingly problematic cases make up the third category as they fall at greater and greater distances from the standard model's ideal (i.e., increasingly close to the right end of the continuum). Correspondingly, the intertheoretic mappings become ever more "bumpy" until, as they near the continuum's opposite end, they become prohibitively so. In this half of the continuum the outlook for reconciling the two theories moves from dim to dismal. New Wave reductionists insist that the failure of intertheoretic mapping in the dismal cases is so thoroughgoing that the success of the reducing theory impugns the integrity of the "reduced" theory and motivates its outright rejection. Ronald Endicott [1998, 57, footnote 13] says that referring to such cases as "bumpy reductions" is like referring to a divorce as a "bumpy marriage."

In these cases that fall at the displacement end of the continuum, the New Wave reductionists' presumptions in favor of neuroscientific over psychological theories has not gone uncontested. Not only have philosophers who are critical of reductionist programs objected, so have many philosophers sympathetic to reductionists' projects — though they have objected *on quite different grounds.* The next

section outlines the two most influential arguments that have arisen from the anti-reductionists. Sections 5, 6 and 7 explore the reservations of other philosophers who are less averse to reductionism, and Sections 8, 9, and 10 trace a confluence of both their and New Wave responses to the two principal arguments of the anti-reductionists that are sketched in section 4.

4 TWO ARGUMENTS FOR NON-REDUCTIVE MATERIALISM

Many philosophers, including many who profess a naturalistic orientation, subscribe to versions of non-reductive materialism that, ultimately, aim to absolve them of much need to scrutinize seriously either reductionist proposals within the psychological and neural sciences or philosophical discussions thereof (whether New Wave or other). These philosophers adopt these positions not because they reject all of the assumptions of New Wave reductionists but rather because they so heartily concur with one of them. Specifically, they agree with the New Wave reductionists' surmise that psychological theories will often show little promise for intertheoretic mappings on to the theories of neuroscience. These non-reductive positions marshal considerations that suggest that reductionists will not be able to map readily either some features of psychological theories or some features of their objects of study (i.e., minds) in to theories about brains in a fashion that will sustain any sort of displacement of the psychological. Their partisans regard one or both of those considerations, viz., the multiple realizability of psychological states or the peculiar character of conscious experience, as establishing barriers to reduction, certainly as classically construed. Instead of employing that premise (in the way that New Wave reductionists do) as promising grounds for *displacing* the psychological, they view it as reason to reject the assumption that such failures to find analogues must automatically impugn the psychological. They hold that, on the contrary, what these failures show is the *indispensability* of the psychological.

Non-reductive materialists come in various stripes but, finally, many take inspiration from what is, by now, a familiar argument concerning the multiple realizability of psychological states.[3] Hilary Putnam [1967/1975] first advanced this argument against psycho-physical identity theories. Putnam argued that the same psychological state can be realized by many different physical states. He appealed to the fact that many organisms other than humans experience pain, yet they have brains that differ considerably from the brains of *homo sapiens.* A physical

[3]Although the brief discussion that follows will address neither his anomalous monism nor his concern with the normative in interpretation, a careful reading of Donald Davidson's "Mental Events" [1970] will disclose assumptions about the range of possible relations between the psychological and the neuroscientific that press an extreme version of the functionalist argument concerning multiple realizability, which is to say that although Davidson avows that every mental event is a physical event, there are no systematic relations to be found among such pairs of event descriptions. Davidson holds that there are no psychological laws; hence there are no *theoretically substantiated* psychological kinds (thus, "multiple realizability" is a bit of a misnomer); hence there are no psycho-physical laws. Of course, Davidson published this paper just as psychology was beginning to free itself from the grip of behaviorism.

description of the arrangements that constitute the hunger of an octopus will almost certainly look quite different from the physical description of the functionally equivalent state in human beings. Therefore, pain, hunger, and, presumably, any of our other psychological states that we attribute to other organisms, cannot be identified with states of the human brain, in particular. It is easy to envision extending this conclusion. Even if we might settle on some state to identify with pain in all of the terrestrial creatures who experience it, that hypothesis would face the further challenge of having to serve as the state of affairs underlying the pains of *extra*terrestrial creatures too. The suggestion is that a wide variety of possible physical states in a wide variety of creatures might all be pain, i.e., that the psychological state of pain almost certainly has multiple physical realizations, and if that is so, then it precludes any simple mapping of this and other qualified psychological states on to the neural states of humans in the way that both psycho-physical identity theories and proposed reductions of psychological theories to theories in neuroscience would require.

The rise of cognitive science and especially of artificial intelligence inspired even more ambitious versions of the multiple realizability argument. As computers proved capable of a growing list of accomplishments from proving theorems to playing chess — accomplishments that we unequivocally regard as intellectual when *we* perform them — it appeared that the chauvinism of neural reductionism and identity theories ran even deeper. Alternative realizations of *bona fide* psychological states were not confined to other critters (including critters from other planets). Completely different forms of hardware could, perhaps, instantiate those psychological states too.

Enter functionalism. Functionalists in the philosophy of mind proposed that the best way to make sense of such a diversity of physical circumstances, all of which could be psychological states, was not to worry about describing these systems physically (since their diversity seemed to guarantee that what was of interest in common about them could not be captured by laws concerning their physical constitutions). They proposed, instead, to characterize psychological states *functionally*. Like most interesting philosophical positions, functionalism, too, comes in many flavors, but it gains traction in debates about reduction when it elevates this thesis to the level of a metaphysical claim [Polger, 2004]. Metaphysical functionalism maintains that as functional states, mental states should be delineated in terms of their causal interactions with one another, with input from the senses, and with motor outputs. On this view, a mental state is the nexus of such causal relations. It is the functionally described operations of a system, rather than any part of the system. It must be characterized at that level of abstractness in order to capture the diverse range of its possible physical instantiations. Employing such abstract accounts of mental states, functionalism, arguably, even allows for instantiations of mental states that are not physical [Fodor, 1981]. Such versions of the position would be non-reductive but examples of neither physicalism nor materialism.

Jerry Fodor [1974 and 1975] appealed to multiple realizability in his criticism of reductionist proposals in psychology. Multiple realizations of psychological states quickly yield reduction functions that are unwieldy and impractical at best. Instead of reduction functions establishing systematic connections between one type of psychological state and one type of neural state within some well-specified set of boundary conditions, the multiple realization of psychological states opens the door to reduction functions that might include immense disjunctions of neural possibilities, since any one of those states of affairs would suffice to instantiate the psychological item in question. Without mappings of psychological types on to neuroscientific types, the multiple realization of psychological states requires that the philosopher of psychology adopt no more than a "token physicalism" that affirms only that each token psychological state is identical with some token brain state. Each token of some psychological type is a token of some physical type, however, every token of that psychological type is not a token of *one* particular physical type. As the possible arrangements that might realize the psychological state increase, so will the size of the corresponding disjunctions.

Fodor contends that disjunctions of possible neural instantiations of some psychological state do not need to be very large before the bridge principles and the reductions that they are taken to inform become not merely unhelpful as guides to scientific research but downright misleading. As noted above, the New Wave reductionists agree, and they take the resulting fragmentation of the psychological categories at the neural level as a reason to expect at least the dismantling, if not the outright elimination, of the psychological account. Fodor, by contrast, stresses that such a displacement of the psychological would result in a science that is needlessly impoverished *from the standpoint of explanation.* It would disassemble perfectly good, readily applicable psychological principles and replace them with a plethora of physical accounts about the micro-level details of diverse systems that would sacrifice all sense of the psychological regularities they exhibit.

The last step of this argument does not turn on anything special about mentality. Fodor's argument suggests that the same morals could apply at any level of analysis in science. He offers an analogy with arrangements within a different level of analysis in science in the service of a *reductio ad absurdum* argument. He compares proposals to replace various concepts in psychological principles with disjunctive summaries of all of their (possible) physical instantiations with proposals to replace the concept 'money' in the principles of economics with disjunctive summaries of all of the instantiations that money has taken in human history. Fodor submits that the former reduction is as pointless and forlorn as the latter (which would truly render economics a dismal science). The absence of anything other than the most multifarious mappings of psychological onto physical states offers no promise of usefully preserving the explanatory achievements of psychology, whether of the folk or scientific varieties. Controversy about whether psychology of either sort, in fact, has many explanatory achievements to be preserved divides participants in these debates along predictable lines [P. M. Churchland, 1989, ch. 1]. Fodor, presumably, thinks that psychology possesses at least enough explana-

tory grip to have motivated this argument in the first place. Even if Fodor's comparative optimism on this count is unjustified and Robert Cummins [2000] is right that most of psychology's principles are not explanatory laws so much as "effects," i.e., patterns in need of explanation, a version of Fodor's argument would still stand. For, whatever the status of psychological principles, the myriad mappings Fodor forecasts would obscure the coherence of patterns in need of explanation no less than that of any putative laws capable of supporting explanations, and in either case the resulting science would have fewer points of empirical leverage rather than more.

The problems with mapping the massive disjunctions that the multiple realizability of psychological states requires do not end there. Those disjunctions also leave the metaphysician at sea with respect to the ontological status of mentality — especially if, as Fodor [1981] entertains, functionalism characterizes mentality at such an abstract level that it might even permit non-material arrangements to instantiate mental states. Without deciding that question, Fodor's anti-reductionism does inspire parody of the early microreductionists' optimism. Highlighting the problems multiple realizability presents for reduction in science, Fodor [1974] advocates the "disunity of science as a working hypothesis."

The bleak reductive prospects Fodor foresees, in effect, insulate psychological theorizing and research. Psychology — and, presumably, any other science in which theories reign whose ontologies diversely map on to the entities that populate theories at lower levels — enjoys great theoretical and methodological leeway, essentially unencumbered by what might appear to be divergent evidence arising from research at lower levels of analysis. Inquiries at those levels merely concern the details of implementation [Fodor and Pylyshyn, 1988]. They will not bear in any integral way on the theories and principles at stake at the higher level. In short, psychology should proceed autonomously. (How all of this fits with Fodor's impatience with "special pleading" in behalf of some sciences is not obvious [1983, 105–106].)

Putnam, Fodor, and the functionalists' arguments have provided fertile ground for metaphysicians eager to preserve the integrity of the mental (and its causal integrity, in particular) without having to surrender their credentials as disciples of modern science. They also seem to permit the promotion of the metaphysics of mentality without risking breaches of materialism. All mental events, states, and (first order) properties are physical events, states, and (first order) properties, respectively. Armed with these multiple realizability arguments, non-reductive materialists can readily acknowledge that brains constitute a material platform on which mentality supervenes, but they need not concede any ground either to identity theorists or to reductionists, who even ponder the possibility of systematic connections between psychological and neural phenomena (typically referred to as "psycho-physical laws"). Finally, the range of possible instantiations and their details also seem to relieve non-reductive materialists from relying on the vagaries surrounding the notions of *emergence* and *supervenience* in order to make their cases. (See [Wimsatt 1997] and Wilson and Craver's "Realization" in this volume.)

Kim, for example, correctly notes that an assertion of "mind-body supervenience *states* the mind-body problem — it is not a solution to it" [1998, 14]. Claims about the emergence or supervenience of the mental are only promissory notes, at best. As with government deficits, the familiarity these notions enjoy and the ease with which contemporary philosophers speak of them does not obviate in the least what is an ever increasing need for their repayment. Non-reductive materialists view functionalism in the philosophy of mind and the multiple realizability argument to which it appeals as a first, major installment against their metaphysical debt.

The second influential anti-reductionist argument looks to the character of our mental experience rather than to technical difficulties associated with disentangling the myriad ways that some psychological phenomenon might be realized by various physical arrangements. This second argument, which is at least as old as Plato's *Phaedo*, holds that it is inconceivable that our conscious experience will ever be explicable or even describable in exclusively physical terms. With the advent of modern neuroscience, philosophers have formulated more finely tuned variations on this position. At a minimum, though, they all deny that any of the conceptions, theory, or language of the sort that inform contemporary physical and biological science will ever suffice to fathom the character of conscious experience. Conscious mental experience may invariably correlate with brain events of specifiable types, but these anti-reductive positions are united in maintaining that no possible amount of information about brains (or any other physical systems) of the type the sciences glean will capture the qualitative character of our conscious experience. These philosophers ([Jackson, 1982; 1986]; [Levine, 1983; 1997]; [Chalmers, 1995; 1996], etc.) have advanced assorted formulations of the features of our conscious experience at issue, but the general point is that even the most comprehensive accounts of the structures and operations of what look to be the relevant brain mechanisms will inevitably fail to convey "what it is like" to see the redness of a ripe tomato, to taste the subtle flavors of a Brussels sprout properly prepared, or to hear Callas' voice.

These anti-reductionists concede that psychological and neuroscientific research may illuminate the comparatively easy problems about mind — basically, those concerning mental contents and access to information, but they insist that these sciences will prove inadequate to the challenges that conscious experience entails. They will not solve the "hard problem" of consciousness [Chalmers, 1995; 1996]. With respect to qualia they will inevitably manifest an "explanatory gap" [Levine, 1983]. The point is not that the qualitative character of consciousness involves features for which we have yet to find any mappings into theories about brains. Rather, their advocates contend, those features are such that no mapping could ever adequately re-describe them. It is this *inevitable* failure to map qualia into theories about brain activities that is the alleged barrier to reduction.

Those non-reductive materialists, who want both the materiality of their cake and their irreducible ability to taste it as they eat it too, may worry that these philosophers' convictions concerning the latter may sabotage their status as materialists. At first blush, the resulting position looks every bit as much like property

dualism as it does like some form of materialism. But if consciousness reflects a level of complexity that is beyond our abilities (or our sciences' abilities) to comprehend, then it is at least possible that there are some physical properties that will ever remain obscure to us [McGinn, 1991]. Given how little of interest we know about our epistemic limitations, the resulting position would seem, at least, to encourage a search for a new fundamental theory of the intrinsic properties of the physical. Owen Flanagan [1992, 128] describes such positions as examples of the "new mysterianism," since from the standpoint of current science these putative intrinsic properties of the physical are mysterious, indeed. Claims about the inability of humans ever to learn more about such properties only introduce an added layer of mystery.

These various non-reductive materialists and New Wave reductionists agree that, whether the issue is *too many* possible connections between psychology and lower level inquiries (the multiple realizability objection) or *too few* (the explanatory gap objection), both constitute barriers to the standard reduction of psychology. Fodor thinks too many connections argue for granting psychology a comparative *autonomy* in its pursuits. Fans of consciousness and the new mysterians, in particular, think that too few (probably zero) connections leave an un-bridged *explanatory gap* and likely point to one that is unbridgeable. New Wave reductionists, by contrast, think that either of these two circumstances will lead at least to psychology's fragmentation [P. S. Churchland, 1983], probably to its dismantling ([P. M. Churchland, 1989]; [Bickle, 1998; 2003]), and perhaps even to its outright *elimination* in at least some of its sub-domains ([P. M. Churchland, 1979; 1989]; [P. S. Churchland, 1986]).

The following three sections take up problems with the New Wave view and with its prognostications about the elimination of psychology. Presenting an alternative account of these matters there and in the remaining sections will include not only richer, more fine-grained analyses of reductive possibilities in psychology and the cognitive sciences that are more faithful to the wide range of considerations that bear on scientific practice but also proposals for defusing both the multiple realizability objection (in section 8) and the explanatory gap objection about consciousness (in section 10).

5 NOT SO NEW WAVE REDUCTIONISM (AFTER ALL): MUTUAL PREOCCUPATION WITH THEORIES

New Wave reductionists have usefully criticized the standard model, exposing the limitations of its restrictive assumptions. The New Wave model's continuum of intertheoretic mapping allows for a range of intermediate relationships that more or less closely approximate the standard model's ideal. They provide little detailed guidance, though, about where approximate reduction collapses into telling discontinuity and about just where on that continuum controversial cases fall. Consider, for example, the putative reduction of Newtonian mechanics to the mechanics of relativity. Nagel [1961/1979, 111] offers this case of intertheoretic relations as an illustration of the standard model of reduction. By contrast, Paul Feyerabend [1962] advances the very same case as a counter-example to the standard model's dictates. However praiseworthy the improvements are that New Wave reductionists have introduced, their analyses also retain at least two *problematic* commitments of the logical empiricists' conception of reduction. They are the first and fourth of four assumptions that anchor the New Wave position.

Standard model and New Wave reductionists assume that distilling sciences down to their *theories* is epistemologically unproblematic. This is the first of the two problematic commitments and the first of the four assumptions. A second assumption, which the New Wave and standard models also share, is that the scientific promise of exploring intertheoretic relations hangs on the set of *functions* (in the most generic sense) that facilitate the characterization of one theory in terms of another. These reductionists hold this view in common, regardless of what conception of theories they prefer, i.e., regardless of whether they construe theories as collections of related statements (logical empiricists) or as sets of mathematical models of characteristic structures [Bickle, 1998; 2003] or as configurations of synaptic weights in brains [Churchland, 1989]. (On this point, see [Endicott, 1998, 62–70].) The third assumption, reviewed in section 3 above, is that, whatever their conceptions of those functions, New Wave reductionists further hold that the relevant collections of those functions connecting particular pairs of theories can be ordered along a *continuum* according to the comparative fidelity of the resulting analogue of the reduced theory. And, finally, however they propose to calibrate and divide that continuum up, partisans for the New Wave model further assume that it provides a *single model of intertheoretic relations with undeviating implications* (depending on where any particular case falls on their continuum) that can account for all of the significant ways in which scientific theories might be connected. This is the second problematic commitment that they hold in common with the defenders of the standard model.

The first half of this section examines grounds for questioning the first assumption, i.e., the mutual commitment of the New Wave and the logical empiricists' models that, ultimately, theories and their relations to one another generally exhaust what is of epistemological interest about cross-scientific connections. In the course of advancing an alternative conception of these matters, sections 6 and 7 present reasons for rejecting the New Wave advocates and the logical empiricists' second mutual commitment that a single interpretation of their models of intertheoretic relations suffices to capture all of the notable intertheoretic connections that arise within science, i.e., the fourth assumption above.

New Wave reductionists often seem to assume, along with their logical empiricist predecessors, that the only epistemologically interesting features (at times they seem to think the only interesting features at all) of cross-scientific relations are the explanatory connections that hold between *theories*. That Paul Churchland's discussions (e.g., [P. M. Churchland, 1989, 48–50]) often exhibit a substantial interest in the reconstruction of a theory's laws is a corollary of this principle. Although New Wave accounts have eschewed the traditional preoccupations with

the *deduction* of the reduced theory's laws, Churchland's equipotent image of the reduced theory constructed within the framework of the reducing theory in an approximate reduction, nonetheless, remains almost exclusively concerned with the reconstruction of the *laws* (and other generalizations) of the reduced theory. That he would show such interest in theories' laws is also somewhat unexpected in light of Churchland's arguments for the advantages of his prototype activation model over the deductive nomological model of explanation. (See [P. M. Churchland, 1989, ch. 10]; [Churchland and Churchland, 1996, 257–264].) By contrast, Bickle [2003, 16] and Schaffner [1992, 329] hold that psychoneural reductions will not involve laws.

This may not be exactly what either the logical empiricists or the New Wave reductionists explicitly say, but it is unquestionably what their discussions of reduction regularly *imply*. Their preoccupation with theories entices New Wave reductionists in their philosophical commentary to minimize the epistemological import not merely of preserving but of cultivating multiple levels of explanation in science and of attempts to integrate the associated inquiries that occur at each level. Instead, their discussions focus overwhelmingly on intertheoretic relations and the putative deflationary implications of intertheoretic reductions for theory and ontology — and for the theories and ontology of psychology, in particular.[4]

They do so even though the neuroscientists and neural network modelers, whose work they discuss, often appeal to findings from higher level psychological sciences to support the neuroscientific and neurocomputational models they prefer. Crucially, those researchers look to these findings (and, by implication, to the methods and techniques by which they were generated) not merely for guidance (e.g., [Hirst and Gazzaniga, 1988, 276, 294, 304–305]) but for *support* for their favorite hypotheses. New Wave and standard reductionists alike concede a role to the special sciences in scientific *discovery*, but the latter, in particular, are clear that the context of discovery is not epistemologically decisive. Both standard and New Wave models' spotlight on the *reduction* of theories in cross-scientific contexts renders them largely insensitive to the contributions higher level sciences regularly make to the *justification* of the very scientific theories — particularly some in neuroscience — that they champion. Neuroscientists and neurocomputational modelers regularly cite psychological evidence in support of their proposals, yet New Wave reductionists often lose sight of all of this in their explicit philosophical analyses and especially in their examinations of the relations between psychology and neuroscience.

Terrence Sejnowski and Charles Rosenberg [1988], for example, appeal to experimental findings in cognitive psychology to support their neurocomputational proposal about the character of human memory. Their famous connectionist model

[4]Bickle [2003, 114 and 130] has conceded an ineliminable "heuristic" role to psychological theorizing and research. Parity of reasoning suggests that theories cast at the sub-cellular and even molecular levels within neuroscience, which Bickle unequivocally prefers, would eventually occupy the same status relative to even lower level theories in related areas of chemistry. (See [Bickle, 2003, 115 and 157]. Contrast [Craver, 2002].)

[1987] for transforming graphemes into phonemes, NETtalk, appears to pronounce written text aloud, once its output is fed into an acoustical synthesizer. Sejnowski and Rosenberg argue that NETtalk's processing and underlying connectionist architecture also offer valuable insights about memory. They advance as a *complementary* theory to prevailing theories in cognitive psychology a new proposal that stresses the *form* of memory representations. Instead of advancing their theory as a competitor that might correct, let alone eliminate, either Bower's encoding variability hypothesis [1972] or Jacoby's processing effort hypothesis [1978] (two hypotheses they explicitly cite), they explore how connectionist architectures could effect such processes and how such connectionist modeling will suggest grounds for the *elaboration* (and, presumably, the reconciliation) of these two cognitive hypotheses.

The first important point here is that the principal evidence for their theory that they cite is the substantial similarities between NETtalk's performance with a particular mnemonic task and the findings in experimental cognitive psychology about the performance of human subjects. As with human subjects, NETtalk's mnemonic performance displays a short-term advantage for massed practice with items and a longer-term advantage for distributed practice, which is to say that it displays the classic spacing effect. This is one of the oldest, best known, and most thoroughly explored findings in the psychology of human memory, dating back to the nineteenth century. It is, for example, discussed in the seminal work in experimental psychology on memory, viz., Hermann Ebbinghaus' *Memory: A Contribution to Experimental Psychology* [1964/1885].

The connections of Sejnowski and Rosenberg's project to research in psychological science does not stop there. One of the familiar experimental designs in memory research in cognitive psychology, viz., cued recall, provides the model for their tests of NETtalk's mnemonic abilities. Since NETtalk is incapable of generating free recalls, for comparative purposes Sejnowski and Rosenberg had to confine their attention to a small subset of the experimental literature on the spacing effect in humans, viz., that concerned with the demonstration of the spacing effect in experiments employing cued recall (e.g., [Glenberg, 1976]). Notably, their argument for the consequence of their findings concerning their neurocomputational model turns, first, on its consilience with findings about the performance of human subjects from experimental cognitive psychology and, second, on the resonances between their tests of NETtalk's performance and the designs employed in experimental work on cued recall [McCauley, 1996].

(For additional examples of the salient role that the integration of evidence arising from various experimental techniques employed in both psychology and neuroscience plays at the borders between these sciences, see [Bechtel and Mundale, 1999]; [Bechtel and McCauley, 1999]; [McCauley and Bechtel, 2001].)

Neuroscientists and neurocomputational modelers regularly cite psychological evidence in support of their proposals, yet the philosophical analyses of New Wave reductionists frequently overlook all of this, especially in their examinations of the relations between psychology and neuroscience. To make this charge is not

to belittle the importance of intertheoretic relations in these settings. Certainly, understanding theories and their relations to one another *is* critical for comprehending much of what goes on in science, however, if the only available accounts of science and the sciences were those of the standard model and its not-so-new-wave successors, a ready inference might well be that *theorizing* exhausts what is epistemologically consequential about scientific activity. The logical structures, material commitments, and mutual relationships of scientific theories are not the only topics relevant either to the justification of those theories or to progressive programs of scientific research. Making sense of theories always requires — in addition to discussions of the theories themselves — discussions of scientific practices and experimental designs, the evidence those practices and experiments generate, and the appraisals of that evidence's import. When that evidence arises from sciences operating at different levels of analysis, it requires at least some attention to the overall structure of science as well. To help clarify some of the issues at stake and to provide an analytical framework for much of what follows will require a sidebar at this point concerning the overall architecture of science and, more specifically, how the various levels of analysis (or explanation) are distinguished.

Talk of analytical or explanatory levels is rampant throughout the literature of the cognitive and neural sciences, but systematic characterizations, let alone precise ones, are rare [Hardcastle, 1996]. A group of criteria for locating levels of explanation among the sciences roughly converge — at least with respect to theorizing about the structural relations of systems. Some of these criteria look to what might be called levels of organization in nature, but ontological and epistemological considerations become rapidly intertwined as these analyses proceed.

Presumably, our most successful theories provide significant clues about the furniture of the universe. This suggests that levels of analysis in science correspond to levels of organization in nature. Typically, what counts as an entity depends on both the redundancy of spatially coincident boundaries for assorted properties and the common fate (under some *causal* description) of the phenomena within those boundaries. For example, both their input and output connections and their various susceptibilities to stains aid in identifying cortical layers in the brain. Emphasizing causal relations insures that explanatory theories in science dominate such deliberations. Wimsatt [1976] suggests that the frequency of items' causal interactions positively correlates with the proximity of the organizational levels at which those items occur. So, usually items at the same level of organization in nature causally interact most often. The greater the number of theoretical quarters from which these ontological distinctions receive empirical support, the less troublesome is the circularity underlying an appeal to *levels of organization* as criteria for their corresponding levels of analysis. Herbert Simon [1969], for example, notes that the amounts of energy necessary to hold systems together increase at progressively lower levels of organization. The forces that sustain the organizational integrity of molecules far exceed those that bind together a block of wood. (This is why experts in karate can break apart blocks of wood, but not molecules, with their hands.)

The *range* of the entities that constitute any science's primary objects of study and its principal units of analysis also offer grounds for distinguishing analytical levels. The lower a science's analytical level, the more widespread the entities it studies. For example, subatomic particles, discussed in physics, are the building blocks of all other physical systems (from atoms, galaxies, and molecules to brains, social groups, and more). By contrast, the minds that psychologists study are only uncontroversially accorded to some (indefinite) subset of biological systems and are parts of socio-cultural systems only.

Such mereological considerations are relevant in distinguishing analytical levels in science but not unqualifiedly so. Analytical levels partially depend upon viewing nature as organized into *parts and wholes*, however the pivotal question for the differentiation of analytical levels is whether or not the wholes are organized or simply aggregates of their parts [Wimsatt 1974; 1986; 1997]. One way of casting the epistemological issue reductionists and anti-reductionists battle about is whether any features of wholes resist explanation in terms of their parts, i.e., whether from an explanatory standpoint wholes are greater than the sums of their parts. The aim is to rule out an account of organizational levels (and, on these criteria, to thereby rule out an account of analytical levels) that tracks simple considerations of scale. For not all big things that have lots of parts (e.g., asteroids or sand dunes) are, comparatively speaking, highly integrated things. If one entity contains others as its parts and if explanations of (some of) its behaviors require appeal to further organizing principles beyond those concerned with those parts, then it occurs at a higher level of organization and likely points to a distinguishable analytical level.

Much of this is an attempt to explicate our intuitions about the *comparative complexity* of systems that arise at different analytical levels. Although complexity has no simple or single measure, we usually have little trouble making comparative judgments about such matters. Cells are more complex systems than crystals. Comparing the relative complexity of systems seems to generate a similar picture of analytical levels in science, with progressively higher levels handling progressively more complex systems.

The order of analytical levels also corresponds to the chronological *order in natural history* of the evolution of systems. Somewhat more roughly, it also corresponds in the history of modern science to the order in which the inquiries in question emerged as disciplines to be differentiated from natural philosophy, as measured by the use of distinctive terms to indicate independent disciplines and the founding of specialized university departments, journals, and professional societies. The lower a science's analytical level, the longer the systems it specializes in have been around. For example, the subatomic particles and atoms that are the principal objects of study in the basic physical sciences appeared quite soon after the Big Bang whereas the compounds and systems on which the biological sciences focus first began to appear (on Earth, at least) somewhere around two billion years ago. Developed nervous systems and brains and the minds that eventually seemed to have accompanied them, by contrast, are more than a billion years newer. And,

finally, cultural systems that the socio-cultural sciences study date from a few million years ago on the very most optimistic estimates and, perhaps, no more than some tens of thousands of years ago (on more exacting criteria). Figure 2 summarizes how this as well as considerations about range and complexity organize the analytical levels of science.

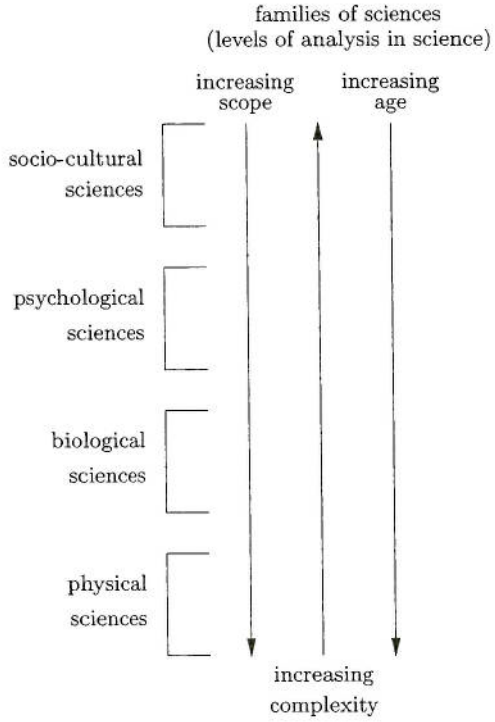

Figure 2.

Methodological considerations also segregate analytical levels but less systematically. Sciences at different analytical levels ask different questions, promote different theories, and employ different tools and methods. Theories at alternative explanatory levels embody disparate idealizations that highlight diverse features of the phenomena on which they concentrate.

The general assumption is that all of these various criteria converge on a grouping of the major scientific families into levels as portrayed in figure 2. Each of these families includes separate sciences that address specific domains; the physical sciences include such sciences as physics and chemistry; the biological sciences include such sciences as molecular genetics and neuroscience, and so on. These sciences, in turn, contain multiple sub-levels [Mundale, 2001]. Employing principles of organization and scale, Churchland and Sejnowski [1992, 11] readily identify seven sub-levels within neuroscience alone (molecules, synapses, neurons, networks, maps, sub-systems, and the central nervous system overall). Figure 3 only begins to hint at the multitude of specialized sciences that have emerged within some of these families of sciences.

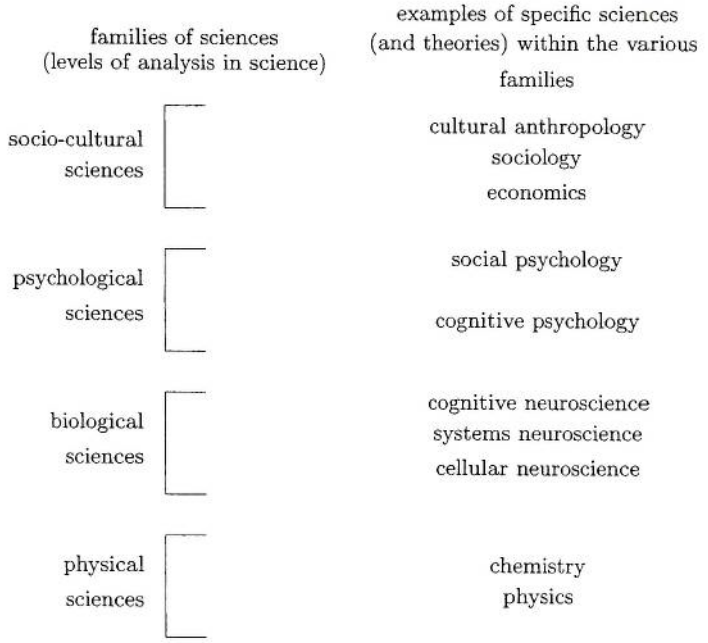

Figure 3.

In contrast to the broad consensus about the groupings of the families of the sciences that figure 2 illustrates, differentiating distinct analytical levels at this more refined register can prove more controversial. This is especially true when trying to sort things out early in some science's history, when decisive accomplishments that end up inaugurating long-standing research traditions that substantially define a sub-level have not yet been recognized as such. The publication of Ulric Neisser's *Cognitive Psychology* [1967] demarcated that sub-discipline within psychology by laying out its salient topics, approaches, theories, and findings and by pinpointing pivotal work by a variety of researchers that in some cases (e.g., [Bartlett, 1932] preceded the appearance of Neisser's book by decades.

6 NOT SO NEW WAVE REDUCTIONISM (AFTER ALL): INSISTENCE ON A SINGLE MODEL

The second problematic assumption that the New Wave reductionists share with the logical empiricists is that a *single model* (or, more accurately, in the New Wave case, a *single interpretation* of their model) can account for all of the (epistemologically) interesting connections between scientific theories. In the same spirit reductionists who subscribed to the standard model also defended a single model of intertheoretic relations, albeit an abbreviated one compared to that of the New Wave or, alternatively, one that assumed (falsely) that every case fell at or quite close to the endpoint of the New Wave continuum designating thoroughly smooth reductions. (Arguably, the logical empiricists' discussions of reduction in science never even countenanced the bumpy cases.) As is the case with the New Wave model, the logical empiricists' single model allegedly describes the critical dynamics connected both with theory succession over time within some science as well as with how the sciences hang together at any particular moment in scientific history, i.e., with intertheoretic relations in cross-scientific contexts.

Clearly, the assumption that a single version of intertheoretic relations (whether the formal standard model or a single interpretation of the New Wave continuum model) can capture the full range of worthwhile scientific connections rests upon their first (shared) problematic assumption that all of the relevant relations are ones or can be condensed to ones between theories. Arguments in the first half of the previous section about the psychological evidence to which researchers regularly appeal in support of their proposals about the structure and operations of the brain (and, thus, that New Wave reductionists regularly presume when they plump for the explanatory promise of some of their favorite candidate reductions) belie that first assumption. The most important consideration here is that making the case in behalf of theories at lower levels quite regularly depends, in part, on appeals to evidence that has been generated in higher level sciences by employing their characteristic methods and experimental techniques.

If these criticisms of the New Wave's first assumption are sound, then they also endanger the current assumption, since, as noted, it depends upon the first. If, as the first assumption asserts, sciences can be distilled down to their dominant theories, then it follows on the New Wave reductionists' account that their comprehensive continuum model of *intertheoretic* relations and their (single) interpretation of its methodological and ontological implications usefully apply to the full range of *cross-scientific* relations as well. This section will offer reasons for thinking that that conclusion is false and, thereby, disclose both why a commitment to a single interpretation of their model of intertheoretic relations is overly ambitious and why any putative reductions of *sciences* (as opposed to theories), including any purported reduction of psychology to neuroscience, are not only not disabling but, instead, a vindication of the "reduced" science and its ontology.

Although their models are less sophisticated than the New Wave models, standard reductionists exhibit some sensitivity to the issues at stake. For example,

even though he too aims to provide a unified treatment of both, in his discussions Nagel [1961/1979] distinguishes between "homogeneous" and "heterogeneous" reductions. Roughly, these options correspond (respectively) to cases in which the reduction functions are straightforward and clear and the terminology is consistent as opposed to those where they are less so. Attention to his discussion and the examples he supplies, however, reveals that Nagel's distinction closely tracks another, more important distinction, viz., one between modeling *progress within a particular science* and modeling *cross-scientific connections*. Nagel discusses the homogeneous reduction of a *theory* and the heterogeneous reduction not only of a theory but also of a *science*. The burden of the discussion that follows is to show that this latter distinction matters (even if Nagel's model mis-describes its implications) and that New Wave reductionists are remiss in ignoring it. Making that case will also uncover grounds for holding that anything anyone is even remotely tempted to describe as "the reduction of a science" will not — contrary to the claims of the New Wave reductionists — jeopardize its importance in the least, let alone lead to its displacement or elimination either in practice or in principle.

Note that *if* sciences could be legitimately distilled down to their theories as both the standard and New Wave models presume, then these putative reductions of *sciences* (say, reductions of perceptual psychology to neuroscience) would also be best understood as reductions of *theories*. The burden of the preceding discussion, however, has been that such a distillation involves unhelpful simplifications and, in particular, unhelpful simplifications about the *justification* of the lower level theories that reductionists prefer.

The homogeneous cases with respect to which Nagel utilizes talk of *theory* reduction concern the relations of successive theories within a single science. These are episodes in the history of science where a new superior theory in some science eclipses what had been the reigning theory. Thus, Nagel employs this language exclusively to explicate, for example, how Newtonian mechanics superseded Galileo's law of free fall [Nagel, 1961/1979, 338–339]. Nagel regards reductions of this sort as relatively unproblematic and comments that they "are commonly accepted as *phases in the normal development of a science*" (p. 339, emphasis added). For Nagel and the logical empiricists homogeneous theory reduction serves as the foundation for their account of scientific *progress*. The new reducing theory and the older reduced theory are continuous with one another logically and materially. Sciences progress through the discovery of new, more encompassing theories that explain everything their predecessors do and more. While the Galilean law only addresses free fall relatively close to the surface of the earth, Newtonian mechanics supplies laws that explain not only this and other terrestrial motions but celestial motions as well. According to the logical empiricists, the history of modern science is a story of the progressive accumulation of discoveries about the world organized by theories employing explanatory principles of increasing generality.

In his most extended treatment of these topics Nagel [1961/1979, 338–339] gives these homogeneous cases of scientific "development" little space. His principal concern is to address what he sees as the more difficult problems that surround

the heterogeneous reductions of *sciences*. The salient point here is how Nagel ends his discussion of the heterogeneous reductions. He chides both allies and opponents of the reduction of sciences for their mutual failure to acknowledge the need to qualify their claims *temporally*. He complains [1961/1979, 364, emphasis added] that " ... questions that at bottom relate to the strategy of research, or to the logical relations between sciences *as constituted at a certain time*, are commonly discussed as if they were about some ultimate and immutable structure of the universe." The aim here is to underscore Nagel's recognition that all such claims concern the various sciences "as constituted at a certain time."

Wimsatt [1976] explicated the critical distinction that lurked behind Nagel's separation of homogeneous and heterogeneous reductions. Nagel's homogeneous reductions of theories dealing with "phases in the normal development of a science" reliably concern what Wimsatt referred to as "*intra*level" relations, i.e., the relations over time between successive theories in some science. The relations of consecutive theories of greatest interest in this context are those between a reigning theory in some science and a competing theory (in the same science) that eventually supersedes it, e.g., in geology from 1950 to 1980, the relations between the theory of the earth's crust as a unitary structure that rises and falls and the theory of tectonic plates or in the neuropsychology of the late nineteenth century, the relations between David Ferrier's [1876] hypothesis that primary visual processing occurs in the angular gyrus and the view of Salomon Henschen [1893], among others, locating it in occipital cortex, instead. Intralevel relations concern theory succession. The crucial point is that these are the relations of competing *theories over time within a particular science operating at a single level of analysis.*

By contrast, Nagel's heterogeneous reductions of sciences pertain to what Wimsatt calls "*inter*level" relations, i.e., the cross-scientific relations between *theories that reign at the same time at different analytical levels in science* — for example, in biology the relations between the theories employed in population and molecular genetics or in the cognitive sciences the relations between psychological conjectures about a specialized capacity for the visual processing of human faces [Farah *et al.*, 1998] and proposals that the fusiform gyrus in the temporal lobe is where such processing is carried out in the brain [Kanwisher *et al.*, 1997].

Wimsatt proposes no longer examining just the logical relations of theories but also the impact of temporal considerations on construing the relations of theories both *within* and *between levels of analysis* in science (see figure 4). Careful scrutiny of these two kinds of contexts will reveal how much their methodological and ontological implications for theories and sciences contrast — especially when the prospects for intertheoretic mappings are bleak [McCauley, 1986; 1996]. Such scrutiny will also provide grounds for resisting the most extreme anti-psychology impulses of New Wave reductionists.[5] Ironically, to see why, it will be useful, first, to register one way in which the New Wave reductionists correctly capture

[5]Such impulses have gradually subsided in the work of some New Wave reductionists. Contrast, for example, P. M. Churchland [1979, esp. ch. 5] with Churchland and Churchland [1996, 219–220] and [1998, 79].

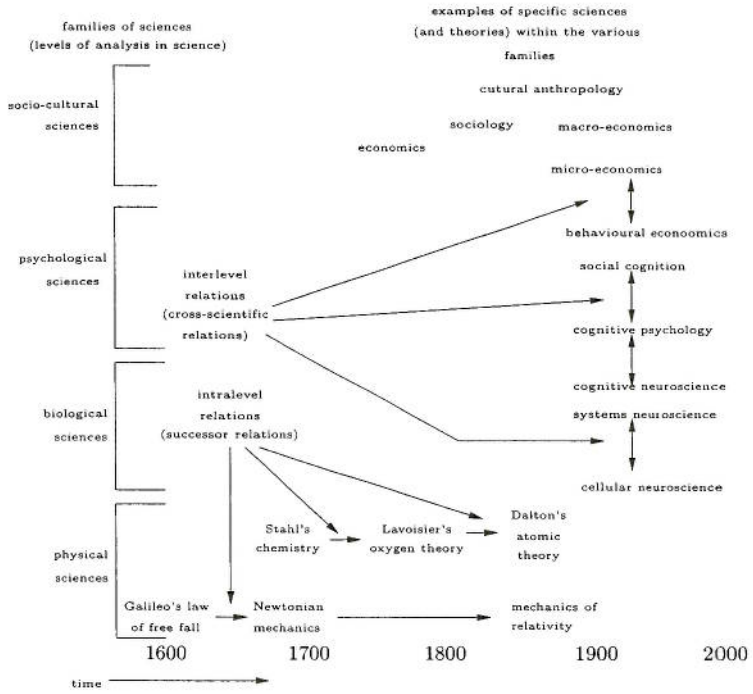

Figure 4.

how these two sorts of cases are *alike*. Both sorts of contexts *do* involve sets of possibilities that range across the New Wave continuum of goodness of intertheoretic mapping. Our ability to construct an analogue of one theory on the basis of a second theory's conceptual resources is unaffected by whether or not those two theories are successors within a single science or prevailing theories employed simultaneously at two different levels of analysis. On this front the two sorts of cases *are* the same.

Note that the point of contention does not concern the applicability of the New Wave continuum in both sorts of intertheoretic settings but rather New Wave partisans' insistence that their *single interpretation* of that continuum's implications will work equally well in both. Contrary to that view, the methodological and ontological implications of falling at some point or other on this continuum can differ substantially in the two different contexts, particularly where the probabilities for finding substantial connections between theories are meager. To help see why, consult figure 5. Introducing the distinction between interlevel (or cross-scientific) relations and intralevel (or successive) relations (portrayed in figure 4) yields *two* contexts in which the continuum of the goodness of intertheoretic mapping applies but, notably, in which its implications diverge when intertheoretic mappings are inauspicious.

Historically, philosophers spotted major cracks in the standard model of reduction when they pondered cases of poor mapping between successive theories in intralevel contexts. As Thomas Kuhn [1970] and Paul Feyerabend [1962] famously emphasized, contrary to the standard model's account of scientific progress, sometimes advances turn not on the discovery of more inclusive theories that are continuous with those that have preceded but rather on largely discontinuous theories, which offer little hope for constructing analogues of their predecessors. Contrary to the standard model of homogeneous reductions, the progress that such revolutionary episodes in the history of science involve does not readily lend itself to characterization in terms of accumulation. In the most extreme cases, the old theories and their ontologies are basically discarded.

New Wave reductionists have learned this lesson well. On their single interpretation of their model of intertheoretic relations the implications of a persisting inability to construct an analogue of one theory within the framework of the other are unvarying regardless of the context — sooner or later, they maintain, one theory's success demands the other's elimination. New Wave reductionists assume that the implications of falling at any particular point on their continuum are unaffected by *how and where* the pertinent theories are situated among the sciences. That contextual, pragmatic, problem solving, and (even) evidential considerations can or should bear on the interpretation of the ontological implications of the varied cases of unpromising intertheoretic mapping never enters the New Wave picture. New Wave advocates presume that any case of serious incommensurability will inevitably and uniformly result in one theory completely superseding another, permanently removing the latter from the scientific stage. In their philosophical proposals at least (see, for example, [Bickle, 1998, 30, figure 2.1]), they

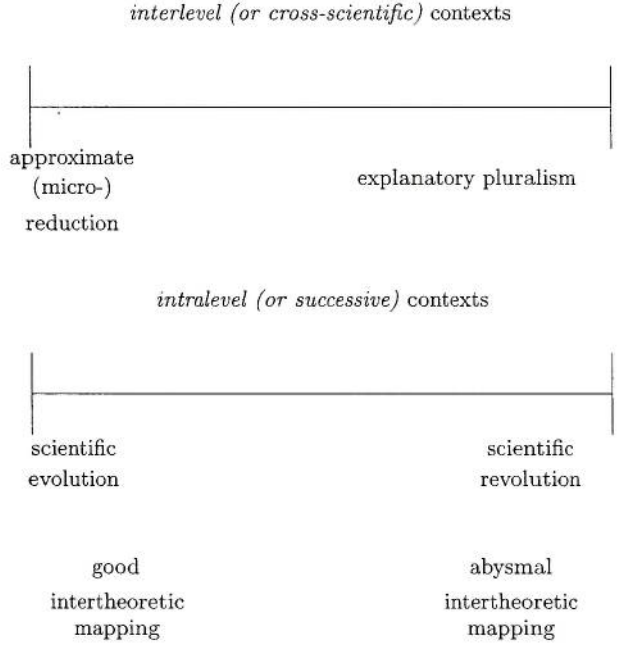

Figure 5.

anticipate this outcome, regardless of

- the amount of empirical support that each theory enjoys,
- the level of explanation in science that each theory occupies,
- the institutional health and longevity of the sciences in which the theories arise,
- the relative status and position of the theories within their respective sciences (for example, are either or both central theories that motivate progressive programs of research?), and
- the amount of fruitful interaction between each theory and other theories at explanatory levels *other* than those at which the two theories in question occur

to mention just some of the more prominent considerations that would influence the probability that one or the other will undergo elimination exclusively on the basis of such a mapping failure.

In fact, these sorts of large-scale, fell-swoop eliminations are an accurate prognosis in intralevel settings only. Where theories are substantially discontinuous in intralevel, successive contexts, the result, when the new theory triumphs, *is* something like a Kuhnian scientific revolution (though, see [Thagard, 1992]). The inability to map out a plausible analogue of the previously reigning theory (e.g., the theory of the bodily humors) employing the conceptual framework of the newly ascendant theory (e.g., modern theories of physiology and of disease) does lead to the elimination of the former theory and some of its accompanying ontology. This is true notwithstanding the persistence of its characteristic idioms in everyday parlance or as modified technical terms in scientific discourse [P. S. Churchland, 2002, 21]. Consider, respectively, the continued use of "choleric" and "sanguine" to describe personality types and the continued use of the term "planet" in the Copernican system or "module" in contemporary psychology [Fodor, 1983, ch. 1].

Reliably, the examples of theory elimination to which New Wave reductionists point — the theories of the crystalline spheres, alchemical essences, phlogiston, caloric fluid, the aether, phrenological faculties, etc. — concern a new theory superseding an older theory at the *same* level of analysis. They concern, in short, *intra*level relations. Because they often fail to distinguish them sharply, New Wave reductionists attribute the profile and outcomes characteristic of intertheoretic relations in intralevel settings to *both* sorts of contexts. Consequently, they anticipate theory elimination when sciences at two different levels of analysis provide disparate views of the same phenomena, i.e., in cross-scientific contexts where the dominant theories at two different analytical levels fail to agree and, therefore, where the prospects for productive intertheoretic mapping (and subsequent co-evolution) do not seem promising. According to the New Wave position, the inability to trace an analogue of the upper level theory within the framework of the lower level theory signals the inevitable elimination of the former by the latter. (Reliably, on the New Wave view it is the upper level theory that should go.) But because New Wave reductionists also subscribe to the (first) assumption that — for most epistemological and ontological purposes — sciences can be distilled down to their theories, the elimination of a theory in these interlevel contexts would be tantamount to the elimination of a *science*! At least some of the time, such going-out-of-business sales are just the result that they predict [P. M. Churchland, 1979; 1989, ch. 1].

The crucial point, however, is that the premier case to which New Wave reductionists wish to apply this moral of the elimination of theory (and, therefore, of a science), viz., the relations between theories in psychology and neuroscience, is precisely a case of *inter*level, cross-scientific relations ([McCauley, 1986]; cf. [Looren de Jong, 1997]). Historical, sociological, and normative considerations argue for why we should neither expect nor wish for the elimination of theories on such grounds in such cross-scientific contexts.

Historical and sociological considerations suggest that once sciences are up and running with research groups, journals, professional societies, university departments and the like — if, for no more reason than social inertia — eliminations of their theories do not occur solely (or even primarily) on the basis of their failure to prove consilient with prevailing theories at other analytical levels. This is particularly so when the putative elimination involves long running sciences and theories from wholly different scientific families (in the case at hand, the psychological as opposed to the biological sciences) that have, in fact, only begun to be very usefully interwoven. That interweaving has arisen as a result in large part of developing better research tools within the higher reaches of recent neuroscience such as the new brain-imaging technologies, PET and fMRI.

Typically, scientists do not look to an adjacent level of analysis in order to shear away its theories and wreak havoc with their ontologies. On the contrary, they are usually looking for help, hoping to find suggestive theoretical, methodological, or evidential resources. To repeat, reductionistic research strategies are probably the single most effective heuristic of discovery in the history of science. Interlevel influence and benefit depend upon forging such connections. In actual scientific practice confronting theoretical incompatibilities across analytical levels does not inspire triumphal campaigns of scientific conquest or usurpation. If anything, it initiates inquiries about points of possible cross-scientific connection. Their development occasions the "co-evolution" of theories at different levels of analysis of the sort that Patricia Churchland [1986; 2002] recommends and, eventually, the emergence of full-blown "interlevel theories" that aid the integration of scientific disciplines. (See [Maull, 1977]; [Darden and Maull, 1977]; [Bechtel, 1986a].) Although *some* of Churchland's commentary (e.g., [1986, 373]) on the probable fate of psychological theories on the basis of their relationships to developments in neuroscience reflect orthodox New Wave views, her discussions of the co-evolution of theories and an "integrationist strategy" [2002, 29] clearly demonstrate her interest in interlevel research.

Normative considerations suggest that this process does not ensue merely because of social inertia. The overall scientific enterprise would be woefully impoverished, if in the face of failures to map theories from two adjacent levels of analysis *at one particular moment in the theoretical evolution of each* (to reiterate Nagel's point) scientists decided to abandon not just the theory but all further investigation at the higher level. Lost would not only be sources of theoretical and conceptual novelty as well as testing procedures and experimental techniques but, frequently, ready access to treasure troves of evidence as well. The many benefits of an *explanatory pluralism* and the multiple analytical perspectives that accompany it and of the co-evolution of theories at different levels of analysis, which ordinarily results, would evaporate, if the failure of intertheoretic mapping in such cross-scientific contexts sufficed for the elimination of one or the other theory and (on the New Wave and standard model's first mutual commitment) the eventual enervation, if not elimination, of the entire science from which that theory springs.

Because the theoretical commitments of psychology (either from the various

areas of experimental psychology or from commonsense psychology) do not thoroughly square with the latest theories and findings in neuroscience is no reason to expect their impending elimination. In addition to the useful role that these notions play not only for practical purposes but in the workings of higher level social sciences, the history of modern science since the mid-nineteenth century provides no compelling precedent for holding either that these theories are prime candidates for elimination or that the psychological sciences, more generally, will likely close up shop (as a result of competition from neuroscience). The following clarification, however, is imperative. To make this claim is *not* to say that either psychological theories or the psychological sciences are autonomous of the prevailing theories (or sciences) at alternative levels (e.g., the neuroscientific) or that they remain in isolation from them or that they are uninfluenced by them. Nor is it to say that they ought to be — quite to the contrary! Eliminations of theory and ontology can occur at any level of analysis in science. The disagreement here is about the origins and the relative power of forces influencing such outcomes.

7 EXPLANATORY PLURALISM IN CROSS-SCIENTIFIC SETTINGS

The Churchlands [1996, 230–231] offer four putative counterexamples to the historical conjecture and normative proposal, offered above, that theory elimination in developed sciences does not (and should not) transpire primarily on the basis of profound theoretical conflicts across readily distinguishable levels of analysis. All four fail.

Responding to their second counterexample introduces no matters of philosophical principle. The putative counterexample concerns the elimination of the theory of caloric fluid by the kinetic theory of heat. The controversy, in this case, turns on the problem of distinguishing levels. In order to make the case that this example concerned an *inter*level setting, the Churchlands press a *controversial* historical claim, viz., that caloric was a *macroscopic* fluid. (This is, of course, in contrast to the uncontroversial claim that its putative action had macroscopically detectable effects). Yet a profound problem for their analysis here is that in her subsequent, detailed discussion of this very case, Patricia Churchland [2002, 21–23] offers compelling historical evidence that caloric was conceived as a *microscopic* fluid! This is the account that squares with the consensus among historians of science, and it undermines their earlier contention that it constitutes a counterexample to the explanatory pluralist's historical conjecture and normative proposal.

The first and the fourth of these alleged counterexamples are particularly startling. They concern theoretical arrangements in seventeenth and eighteenth century biology and in atomic chemistry respectively. The Churchlands note, in the first case, that "the conceptual framework of early biology ... was eventually displaced by an entirely new framework of biological notions ... regularly *inspired* by the emerging categories of structural and dynamical chemistry ... " and in the fourth that " ... most of the details of Dalton's atomism ... were *inspired* by higher-level chemical data concerning ... constant weight ratios experimentally revealed ... " [1996, 230–31, emphasis added]. (They go on to stress, quite correctly, that in the latter case the cross-scientific influence was *top-down*, not bottom-up.)

These alleged counter-examples startle, because they so clearly miss their target. These are counter-examples to the claim that prevailing theories in sciences operating at different levels of analysis remain in complete isolation from one another. Unfortunately, no one (not even Fodor) holds that view. To repeat the culminating claims of the previous section, to resist assertions about the elimination of whole theories and sciences in cross-scientific contexts is *not* to deny cross-scientific influences. On the contrary, a critical feature of explanatory pluralism is to highlight the prominence of such influences. After all, the benefits of the co-evolution of theories in science will not arise, if theoretical incongruities in interlevel contexts invariably demand either the elimination of theories and sciences, on the one hand, or — just as unhelpfully — their thorough or perpetual autonomy, on the other.

Explanatory pluralism *underscores* the on-going interaction of scientific enterprises carried out at the various analytical levels. All scientific explanation is partial explanation from the perspective of some analytical level or other. Scientific theories and the explanations they inspire are selective. They neither explain everything nor wholly explain anything (cf. [Polger, 2004, 203]. Explanatory pluralism denies that inter*theoretic* relations exhaust all of the selection pressures in the resulting co-evolutionary process and that those selection pressures are exerted exclusively from the bottom up (as the Churchlands' treatment of Dalton's atomism correctly headlines). Bickle comments, for example, that even the parade-case reduction of classical thermodynamics to statistical mechanics involved "mutual feedback" [2003, 11]. So, when the Churchlands tout the abilities of concepts and theories at lower levels (in the first case) or findings at higher levels (in the fourth case) to "inspire" developments at other analytical levels, what they have provided are not counterexamples but *illustrations* of the explanatory pluralist's account of cross-scientific dynamics that they take themselves to oppose. The plurality of explanations at multiple analytical levels is a necessary precondition for the "integrationist strategy" Patricia Churchland [2002, 29] advocates, in which developments (not just theoretical developments, incidentally) at various levels of analysis, both within and between the families of the sciences, inspire and enrich inquiries at alternative levels concerned with the same phenomena under descriptions of different grains. In short, these two of the Churchlands' putative counter-examples are utterly unconvincing.

The Churchlands' third alleged counterexample brings things full circle, since it explicitly addresses the *reduction of a science*. They [1996, 230] maintain that by reshaping our understanding of what light is Maxwell's electromagnetic theory showed:

> ... the well-established conceptual framework of geometrical optics, while a useful tool for understanding many macro-level effects ... to be a false model of reality when it turned out that all optical phenomena could be reduced to (i.e., reconstructed in terms of) the propagation of

> oscillating electromagnetic fields. In particular, it turned out that there is no such thing as a literal *light ray*. Geometrical optics had long been inadequate to diffraction, interference, and polarization effects anyway, but it took Maxwell's much more general electromagnetic theory to retire it permanently as anything more than an occasionally convenient tool.

They offer this episode in the history of science as a putative counterexample that will "contradict" the claim that *eliminations* of higher level theories and their ontologies in cross-scientific settings are not driven exclusively (or even predominantly) by incompatibilities with theories at lower levels [Churchland and Churchland, 1996, 224]. The case of geometrical optics fails to serve that purpose, however, because *the interlevel mapping here is comparatively good* (as the Churchlands themselves note). Consequently, it is the wrong sort of case. This case approximates the microreductive ideal. It falls on the *left* half of the upper continuum in figure 5, not the right! Nor does this case involve any fell-swoop eliminations of either theory or ontology. On the contrary, any putative *reductions of a science*, i.e., reductions in situations that involve *good* intertheoretic mappings in interlevel contexts, lead neither to the elimination or the displacement of an upper level theory, but rather to its *vindication*.

It will help to begin by pointing out a consideration that will *not* be employed in making that case. The Churchlands explicitly note that geometrical optics remains "an occasionally convenient tool," presumably for the purposes of rough calculation. (In light of its pervasive use in everyday contexts, to describe the convenience of geometrical optics as "occasionally" useful only may be a bit of an understatement.) Long abandoned theories are, however, often employed for their convenience in calculation. The principles and tools of traditional celestial navigation, for example, enable mariners to calculate their positions, yet they presume both geocentric and geostatic arrangements. So, the fact that geometrical optics remains the standard conceptual tool for many everyday calculations at the macro-level does *not* demonstrate that the theory and its ontology have not been eliminated from the canonical commitments of science.

The Churchlands' comments that "there is no such thing as a literal *light ray*" and that geometrical optics offers a "false model of reality" (and, of course, the fact that they offer this case as a counterexample in the first place) signal the candidacy of geometrical optics, in their view, for just that sort of elimination. Yet everything about their description of this case suggests that they finally regard the reduction of geometrical optics to electromagnetic theory as a prototypical case of a New Wave *approximate (micro-)reduction* of a technically false theory (in the same sense that Mendelian genetics or Newtonian mechanics are technically false theories). What the Churchlands show is that geometrical optics provides an obviously *incomplete* account of optical phenomena, not that it is glaringly false in the way that we regard, say, the phlogiston theory of combustion to be.

To repeat, the point about cases of approximate (micro-)reduction is expressly that the intertheoretic mappings they involve are *comparatively good*. The Churchlands, after all, emphasize the ability of electromagnetic theory to reduce, i.e., reconstruct in its terms, "*all* [of the relevant] optical phenomena" that geometrical optics surveys — including, incidentally, precisely the phenomena still commonly referred to as "light rays" ([Churchland and Churchland, 1996, 230], emphasis added). There may not be any such thing as a literal light ray, but, quite *unlike* phlogiston, for example, there certainly is some thing to which the term "light ray" refers, viz., a direction perpendicular to that of the relevant waves of electromagnetic radiation.

According to the Churchlands' own description, then, this case falls in the *left* half of the *upper* continuum in figure 5. Apparently, like the classical gas laws, the principles of traditional geometrical optics are unable to match the precision and scope that the corresponding lower level theory provides. Also like the classical gas laws, though, they constitute "a useful tool for *understanding* many *macro*-level effects" ([Churchland and Churchland, 1996, 230], emphasis added). Arguably, conceptual tools that create an *understanding* of effects provide deeper, more valuable, cognitive insights than those which merely predict those effects or facilitate calculations. (This claim is a corollary of the prototype-activation model of explanatory understanding that Paul Churchland defends. See [P. M. Churchland, 1989, ch. 10]; [Churchland and Churchland, 1996, 257–258 and 264].) That geometrical optics manages to do all three testifies to its probity. The salient point here, though, is that *its approximate microreduction* to electromagnetic theory *adds* to that testimony.

Scientists' ability to construct relatively faithful analogues of the principles of traditional geometrical optics within the conceptual framework of electromagnetic theory amounts to even more compelling evidence of its integrity. That traditional geometrical optics — so far as it goes — basically squares with electromagnetic theory is an asset, not a liability. The heightened precision and generality and the added insights (concerning just such things as diffraction, interference, and polarization) in this domain and others that electromagnetic theory brings do not discredit the accomplishments of traditional geometrical optics. Nor did they bring research in geometrical optics, transformed and enhanced by its reductive integration into electromagnetic theory, to a halt. Consider the work of Lord Rayleigh and, for example, the computation of the Rayleigh limit.

Again, like the classical gas laws and Mendelian genetics and *quite unlike the theory of the bodily humors or the chemistry of Stahl*, the basic principles of geometrical optics are construed as broadly continuous with subsequent theory and regularly serve in educational programs for mastering the modern science. Not only does mapping an upper level theory comparatively well on to a successful lower level theory not impugn it, it vindicates it. That such upper level theories can serve as heuristics of calculation settles questions about their *practical* value straight away. To the extent that an upper level theory either has, in the past, provided sound principles that organize phenomenal patterns and offered explanatory insights that are *corroborated* by its reductive relationship with a more inclusive lower level theory or, in the present, continues to offer theoretical guidance and in-

spiration, productive experimental techniques, and bodies of evidence (to which, among others, theoreticians working at lower levels may appeal) establishes its fundamental contribution to the scientific enterprise and should discourage talk of its elimination. That some of these observations are generally of a piece with many of the Churchlands' comments at other points about such cases of approximate reduction (e.g., [P. M. Churchland, 1989, 47–50, 215]) renders their appeal to the case of geometrical optics as a putative counter-example to the explanatory pluralist's claim that the elimination of theories does not occur as the result of interlevel conflicts all the more puzzling.

Endicott argues that New Wave reductionism faces a dilemma with just such cases of approximate (micro-)reduction, since once the co-evolution of theories between levels begins, the notion of independently constructing an analogue of the reduced theory (within the framework of the reducing theory) on which that reduced theory has had no influence is either implausible or will look superficial and artificially abstemious from a historical standpoint. About cases just like the microreduction of geometrical optics, Endicott [1998, 67] comments "on the worst case scenario, new-wave construction is flatly contradicted by co-evolutionary facts; on the best case scenario, it is historically shallow and methodologically restrictive."

Fears about a *science's* elimination or enervation or dispensability on the basis of its alleged (standard or New Wave) reduction are unfounded. Reductions of psychological theories to theories in neuroscience are conceptual bridges that permit intellectual traffic to flow between these two sciences. Constructing such integrative relationships amounts to establishing the sort of infrastructure that facilitates important forms of scientific progress at both levels. Nor does endorsing either the power of reductive strategies in science generally or individual reductive proposals at the border of psychology and neuroscience specifically require any expectations about sciences that operate at higher analytical levels eventually being forced to conduct going-out-of-business sales. It never has been nor is there any reason to think that it ever will be the case that up-and-running sciences collapse on the basis of successful theoretical reductions, or that they could so collapse, or, especially, that they *should* so collapse. *Sciences* are not the sorts of things that get reduced or eliminated. Theories are.

When considering reductive relations between theories in psychology and neuroscience, it is critical to realize that the arguments that have been offered above for those last claims are not the familiar ones that either Anglo-American or Continental philosophers, who fashion themselves friends of subjectivity or intentionality or consciousness or individuality or contextuality or human freedom or the great tradition of humanism, propound. Contemporary naturalists, who are interested in the cross-scientific connections between the psychological and cognitive sciences, on the one hand, and neuroscience, on the other, are not attracted to bare philosophical conjecture in the domain of mind and brain. The Churchlands [1998] defuse five well-worn objections to the reduction of psychology to neurobiology that appeal, respectively, to sensory qualia, intentionality, complexity, freedom, and multiple realizability. Bickle [1998] employs essentially the same strategy that the Churchlands do to dismantle the multiple realizability objection. Bickle also concurs with Kim [1989] that boosters of supervenience and non-reductive materialism have, unfortunately, misunderstood the inevitable property dualism their positions entail. Bryon Cunningham [2001a; 2001b] provides unified replies to anti-reductionist arguments that look to complexity and qualia, respectively.

The last three sections examine the confluences of views among naturalistically oriented philosophers interested in cross-scientific relationships where interlevel connections seem to promise a co-evolutionary integration of sciences and the eventual development of interlevel and interfield theories [Maull, 1977; Darden and Maull, 1977]. Like the emergence in the middle of the twentieth century of biochemistry at the border of two of the major families of sciences, the blossoming of cognitive neuroscience over the past few decades has signaled the development of an interlevel science at the border between the biological and psychological sciences. The development of brain imaging technologies has only abetted the prospects for even richer integration.

The Churchlands' comments on geometrical optics to the contrary notwithstanding, much of the time New Wave reductionists seem to agree with explanatory pluralists (and those whose work exhibits the morals of explanatory pluralism in the study of complex mechanisms and mechanistic explanation in science). They concur about both the prospects for and the implications of such scientific integration in the areas of the psychology-neuroscience interface that fall in the left half of the upper continuum in figure 5. At the very least, none of these philosophers subscribe to either the multiple realizability objection or the explanatory gap objection to the reductive integration of the psychological and neuroscientific. Their responses to these objections in large part take their inspiration from patterns in the history of research in science and in cognitive neuroscience, in particular.

8 MULTIPLE REALIZABILITY AS AN ARGUMENT FOR (NOT AGAINST) REDUCIBILITY

It is precisely concerning cross-scientific relations where sciences at adjoining levels of analysis offer fruitful explanatory proposals, where interlevel conflicts are neither plentiful nor weighty, and, thus, where intertheoretic mapping is promising and where the co-evolution of theories *and integrative empirical research* are underway (for example, in cognitive neuroscience) that contemporary accounts of reductive integration (whether New Wave or explanatory pluralist) most resonate with one another. In his most recent work even Bickle, "ruthlessly reductive" by self-description, who has argued most steadfastly for increasing deference to lower level theories and explanations as a principle that organizes scientific progress, now lauds the co-evolution of psychology and neuroscience, eschews the elimination of psychology, and recognizes its vital contribution to the characterization of higher cognitive functions ([1998, 141]; [2003, 128 and 130]). The complicated brain activation patterns that brain imaging technologies reveal would, for ex-

ample, be virtually meaningless without theoretical and methodological guidance from psychology [Mundale and Bechtel, 1996]. Bickle also allows that progress in neuroscience, at least, requires methodical research at "multiple levels" and that "psychological causal explanations still play *important heuristic roles* in generating and testing neurobiological hypotheses"[6] [2003, 178 and 114]; cf. [Craver, 2002]).

Because of their attention, first, to particular processes and mechanisms that both psychology and neuroscience address and, second, to the lessons from the history of research and practice in these and other areas of science, virtually all contemporary philosophers, who are at all sympathetic to reductionist projects, remain unimpressed by the two major philosophical objections to the reductive integration of psychology and neuroscience. Among the things that these naturalists hold in common is that the relevant sciences here have progressed to the point where, like the philosophies of physics and biology before them, the philosophies of psychology and neuroscience can no longer afford to prize philosophical cleverness or metaphysical comfort over empirical accountability and explanatory adequacy.

Initially, the two major replies to the multiple realizability objection might seem antithetical, since the first headlines how *often* multiple realizability arises in science, while the second focuses on how *infrequently* it arises at the psychology-neuroscience interface, at least once researchers are clear about the best models available in each science and about how coarse or fine-grained those models are. In fact, the two replies are closely related, with the second reply building on the first. The first reply denies that the anti-reductive conclusion follows from a premise affirming multiple realization. The second reply suggests that the range of *theoretically interesting* realizations may have been greatly exaggerated and that what few exist will prove eminently manageable.

[6]On the other hand, it would be easy to exaggerate the resonances between the various philosophical conceptions of cross-scientific relations that hold out promise of extensive integration.

Like the Churchlands' take on geometrical optics, Bickle sometimes seems to lose sight of the support an upper level theory enjoys in an approximate microreduction and of the contributions it can still make to on-going inquiries when he holds that model building ceases at the psychological level and that psychological models lose their causal explanatory status once reductive integration has commenced [2003, 110–111, 114]. So, Bickle argues that in the face of the reduction of the consolidation of declarative memories to the operations of the cellular and molecular mechanisms that drive the shift from E-LTP to L-LTP and that selectively preserve L-LTP in certain synapses [Lynch, 2000], "it seems silly to count psychology's "explanation" of consolidation as "causally explanatory," "mechanistic," or a viable part of any *current* scientific investigation *still worth pursuing*" [2003, 112].

Maurice Schouten and Huib Looren de Jong [1999] argue that such claims misread the on-going contributions to the specification of the underlying mechanisms of functional analyses in psychology (which can, among other things, implicate relations with socio-cultural phenomena external to the organism). They maintain, in effect, that downplaying psychological considerations in this way strikingly misjudges just how long the co-evolution of theories and entire sciences can go on, which, presumably, arises here from an underestimate of the range of psychological questions that are of interest, for example, concerning memory. Rather than worrying about settling attributions of causal responsibility, Schouten and Looren de Jong (like [Bickle, 2003, 31–40] advocate attending to the details of particular mechanisms. They argue that Bickle overplays the import of findings concerning LTP for the reduction of the wide range of findings and patterns that constitute the psychology of memory. (See [Bickle, 2003, 45–46].)

The first reply highlights the *pervasiveness* of lower level multiple realization of higher level phenomena throughout the sciences and, particularly, in cross-scientific contexts that involve uncontroversial reductions. Multiple realizability often arises even in the parade cases of the standard model, such as the reduction of the notion of temperature in classical thermodynamics to an account in terms of the kinetic theory of heat. Technically, though, what the kinetic theory readily reconstructs is temperature-in-a-gas. The Churchlands [1998, 78] note that "in a gas, temperature is one thing; in a solid, temperature is another thing; in a plasma, it is a third; in a vacuum, a fourth; and so on ... this ... just teaches us that there is more than one way in which energy can be manifested at the microphysical level." As Robert Richardson [1979; 1982] has emphasized, reductions in science are generally *domain specific*, and Nagel's comments on the character of the bridge principles linking the predicates of the reduced and reducing theories indicate that he already recognized this.

The goal of this first argument is to domesticate multiple realization. Philosophers should be less impressed with the multiple realizability objection once they appreciate the frequency with which multiple realizability occurs in nature. Although he confines his analyses to the logical empiricists' standard model of reduction, Kim [1989, 39] claims that the domain specific reductions, which result under such circumstances, abound throughout the sciences and not merely in psychology and that they are "reductions enough ... by any reasonable scientific standard and in their philosophical implications."

Still, might the anti-reductionist reply that the multiple realizability of *psychological* states is of a wholly different order? After all, what about Fodor's arguments that disjunctions of possible realizations of psychological states at the physical level need not be very large before they become unmanageable from a practical standpoint and unhelpful from the standpoint of explanation? The anti-reductionist can acknowledge widespread multiple realization in many other cross-scientific contexts but argue that the psychological case brings added problems all of its own.

The most troublesome problems here do not concern the psychologies of aliens or robots, since for many purposes the division of psychology and cognitive science into specialized sub-domains seems plausibly motivated on a variety of criteria in the same way that accounts of heat in gases, solids, plasmas, vacuums, and so on are usefully distinguished for some of our problem solving purposes in physical science [Mundale and Bechtel, 1996, 490]. Not even the multiplicity of non-human organisms' psychological states (recall the hunger of octopuses) presents the biggest challenge here. There may be fewer grounds for sub-dividing psychology by phylum or species in the light of the commonalities of their evolutionary heritages, but both the physical and the (probable) psychological differences look substantial enough in most cases that even this sort of specialization in psychology does not look utterly unmotivated. These forms of multiple realizability present challenges, but they are not insurmountable challenges, since the sub-division of some psychological concepts for some explanatory purposes is no more out-of-bounds than

is the occasional sub-division of the concept of heat in thermodynamics that the Churchlands raised.

The multiple realizability that Fodor's analysis points to is of a more fundamental sort. Sub-dividing psychology along the lines of basic material substrates or according to distinctions among species would circumvent many of the problematic disjuncts that would make for unwieldiness in the bridge principles of a reduction, but its sub-division between individuals or worse yet between the same individual at different times would drain away any possible interest in the putative reduction.[7] As William Bechtel and Jennifer Mundale [1999, 177] note, " ... even within a species brains differ. Even within an individual over time there are differences (neurons die, connections are lost, etc.). Thus, multiple realizability seems to arise within species (including our own) and even within individuals." The rarefied sub-divisions within a reduced psychology that these forms of multiple realization portend surely expose the futility of the reductionist's project. Apparently, psychological insights would be inundated and sink, lying lost along the bottom of vast oceans of neuroscientific detail.

Bechtel and Mundale, however, do not expound on this objection concerning the proliferation of possible neural instantiations of human psychological states in order to praise it, but rather to bury it. Multiple realizability arguments look plausible, first, because anti-reductionists have generally failed to attend to what scientists have ascertained to be the theoretically significant kinds at each analytical level (and especially at the level of neuroscience) and, second, because they have also consistently ignored whether the kinds they do discuss are cast at comparable grains.

On the first front, Bechtel and Mundale [1999, 203] remark that, as Leibniz observed three centuries ago, any two particulars will both resemble and differ from one another in an infinite number of respects, and there is no reason to expect things to be any different when comparing brain states and psychological states. Science is always concerned with ascertaining which resemblances and differences *matter* from the standpoints of explanation, prediction, and control. The aim is not to map each and every homespun category we may employ in these or any other domains, but rather to concentrate on those that our best explanatory theories spotlight [Hardcastle, 1996].

Such considerations may even partially neutralize the sting of Fodor's famous argument about the fruitlessness — for understanding economics — of a focus on the various instantiations of money. Attention to the limitations that particular material forms that money can take impose will disclose some eminently useful, though admittedly low level, generalizations about those forms' deployment within economies. For example, some transactions such as mortgage closings at banks and purchases of items stored in inside pockets of less scrupulous vendors' trench coats in alleys in large cities will almost never involve personal checks or credit cards or, at least at the mortgage closings, large amounts of cash. Thus, some

[7] Davidson [1970] employs different premises but the conclusion he draws, viz., that there are no psycho-physical laws, is functionally equivalent.

patterns in the economic domain may offer grounds for the fragmentation of the concept 'money' along these lines for certain limited, domain specific explanatory purposes.

On the second front, philosophers find ubiquitous multiple realizability in psychology because they regularly compare coarse grained psychological concepts with exceedingly fine-grained conceptions of brain states. The folk psychological notions that particularly interest philosophers are more coarse grained than most employed in experimental cognitive psychology, while the conceptions of brain states they discuss, Bechtel and Mundale argue, are much finer-grained than the ones practicing neuroscientists use in their theories. They comment [1999, 178] that "when a common grain size is insisted on, as it is in scientific practice, the plausibility of multiple realizability evaporates." The point is not that any particular grain is canonical, but only that, once philosophers compare psychological and neuroscientific blueprints of comparable scale, multiple realizability vanishes.

When Putnam [1967] examined hunger in humans and octopuses, his grain for type identifying psychological states was not especially fine. Certainly, such a broad extension of psychological types poses problems for the *functional* identification of psychological states, since the links to other mental states and to behaviors that are central to functional analyses differ profoundly between such radically different species. (For example, as the result of their hunger, octopuses never ponder a quick trip to the supermarket nor do they ever slap some peanut butter and jelly on some bread for a quick snack.) Still, given that evolution tends to conserve and extend existing mechanisms rather than create new ones, researchers could well end up type identifying even the neural mechanisms involved in hunger in the octopus and human, which would substantially defuse Putnam's intuitively plausible example. This is not to rule out the possibility of radically different ways of performing similar functions emerging in evolution. However, the point is precisely that when researchers discover evidence of multiple mechanisms for performing similar functions, such as alternative pathways for processing visual input in invertebrates and vertebrates, it provides an impetus for psychologists to search for functional (behavioral) differences that motivate the differentiation of types at the psychological level as well (cf. [Kim, 1972]).

Ascertaining compatible grains between research at two different levels is one of the most basic steps in the co-evolution of sciences. Getting the grains right between theoretically significant kinds makes all the difference. A variety of *successful research strategies* at the borderline of psychology and neuroscience, some of which have, by now, been *utilized for more than a century*, tacitly repudiate the multiple realization of theoretically relevant psychological states. This is not only true about the interpretation and the integration with models in cognitive neuroscience and cognitive psychology of recent findings from PET and fMRI research, where scientists have obtained generalizable results by employing sophisticated statistical techniques to analyze multiple images of multiple brains. Neuroscientists' inferences about the cognitive functions of various areas in unimpaired brains, on the basis of studies, such as Paul Broca's ([1861]– cited in [Bechtel and Mundale, 1999,

184]) classic work, on performance deficits and brain damage, have simultaneously proceeded unencumbered by worries about multiple realizability and proven one of the most fertile research strategies available.

The force of the first reply to the multiple realizability objection was to regularize it by stressing how often it arises in science. The force of this second reply, which explores the consequences of the first for scientific practice, is to accentuate how rarely multiple realizability presents any barrier to reduction in science, at least once we get clear about both the operative explanatory categories and their comparative grains.

Bechtel and Robert McCauley [1999] push this second reply one crucial step further. They argue for a conclusion that, if sound, not only deflates the anti-reductionists' multiple realizability objection but construes multiple realization as a platform for *defending* a version of the position that the objection was formulated to waylay! Bechtel and McCauley point out that for a host of reasons, beginning with ethical ones, the overwhelming majority of the research and experimentation in the history of neuroscience has been done on the brains of non-human animals. This is to say that neuroscientific research on the identification of brain areas and processes is done *comparatively*.

Korbinian Brodmann [1909/1994] used a variety of criteria to map the human brain into functionally distinguishable areas. These included attention to the gross anatomical features of brains as well as the examination of cortical micro-structure. Brodmann employed cytoarchitectural tools to demonstrate that cortex generally consists of six layers. He distinguished brain areas, in part, on the basis of the relative thickness of these layers (e.g., layer 4 was very thick in areas 1, 2, and 3, but much thinner in area 4) and the particular types of neurons (e.g., pyramidal cells) found in them. The important point is that Brodmann based his account of cortical layers on *comparative studies involving fifty-five species.* Brodmann also proceeded comparatively in his study of neuroanatomical features. In addition to his well-known map of the human cortex (figure 6), Brodmann produced maps for an assortment of other species, including the lemur, flying fox, rabbit, hedgehog, and others.

For Brodmann finding comparable areas in different species despite differences in brain shapes and in the relative location of areas was pivotal in identifying and building a case for functionally distinct areas in the human brain. In his work and that of other neuroscientists concerned with such questions, the multiple realization of some psychological function across species in related but different structures does not obstruct the identification of an area. On the contrary, it is the single most compelling type of evidence available for identifying an area in the human brain! Contrary to contemporary anti-reductionist orthodoxy, multiple realization across species is not a barrier to the mapping of some psychological function on to brains, rather it is the key to accomplishing such mappings.

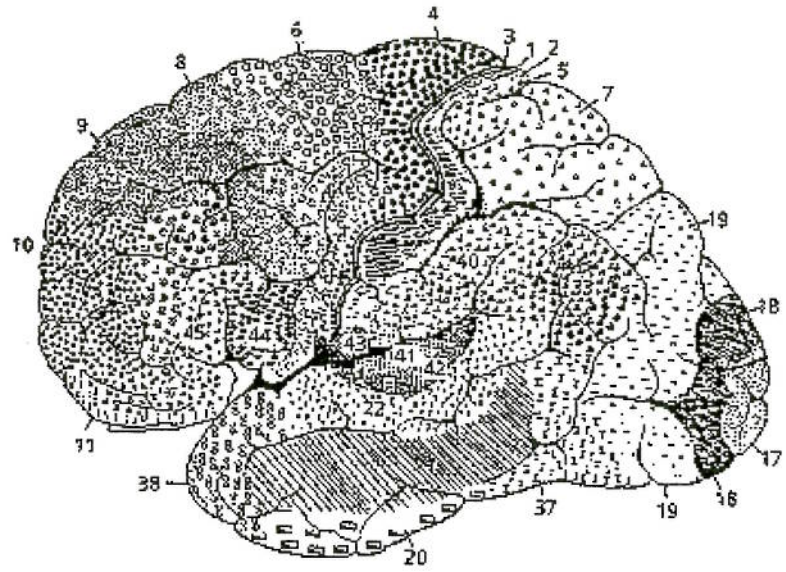

Figure 6.

9 MECHANISTIC ANALYSIS AS EXPLANATORY PLURALISM WRIT SMALL

Recasting multiple realizability as an aid, rather than a barrier, to the integrated development of the psychological and neuroscientific has encouraged some naturalistically oriented metaphysicians to reassess the relative promise of functionalism and the psycho-physical identity theory. This new perspective on multiple realizability suggests that even if the premises that inspire functionalism are true, they block neither the psycho-physical identity theory nor the reductive integration of a good deal of the psychological with the neuroscientific.

Thomas Polger [2004] accepts Bechtel and Mundale's argument about the decisive importance of ascertaining the grain at which various psychological and neuroscientific descriptions are cast when pondering the merits of the identity theory. A perfectly interesting version of the identity theory need only require that humans' mental states are identical to some of their brain states. It does not require that all mental states, especially those cast at coarse grains within psychology, are identical to one another or that all identities of psychological and neural states are cast at a single grain. To repeat, the point is *not* that some grain or other is canonical, but rather that preferences concerning grain in any given case have ev-

erything to do with added empirical accountability, explanatory accomplishment, and promise of productive extensions of research and of cross-scientific consilience — in just about that order of decreasing significance.

Polger argues that, finally, the only version of functionalism that matters for these controversies is one that maintains that functionally distinguishable mental kinds merit an independent metaphysical status. According to Polger [2004, 136], abstractness of its functionally discriminated kinds, relative to the categories deployed in the explanatory theories of biological science, is the critical condition such a version of functionalism must satisfy in order to outstrip the identity theory. He argues that none of the prominent accounts of the functional advanced in the literature can legitimately purchase that biological abstractness without sacrificing one or more of three other necessary conditions for a satisfactory conception that he specifies, viz., causal efficaciousness, objectivity, and synchronicity [2004, ch. 5, esp.177–178].

Polger's proposal [2004, 188] for resolving the complications, which the possibility of multiple realizations of psychological items introduces and which originally motivated functionalism in the philosophy of mind, echoes Bechtel and McCauley's [1999] argument. As with any other cross-scientific case, the conceptual and theoretical resources of psychology and neuroscience provide a variety of different sub-levels and associated grains at which the relevant events, states, structures, systems, and processes may be cast. Identity theorists have every right to search for patterns and mechanisms at different levels and grains on a case by case basis. That is in the service of sub-dividing these phenomena according to the lights of the best explanatory theories available. Polger holds that the considerations that most philosophers seem to think point toward functionalism will, in fact, just as readily square with a version of the psycho-physical identity theory. What he, ultimately, seems to anticipate is a theory of mind that is informed by the lessons of the explanatory pluralist's picture of cross-scientific relations.

Polger is clear that any version of the identity theory he envisions is not one that turns on consummating either standard or New Wave reductions.[8] He turns for inspiration, instead, to the recent literature in the philosophy of science on mechanisms and mechanistic explanation. (See [Bechtel and Richardson, 1993]; [Machamer, Darden, and Craver, 2000]; [Craver, 2001]; [Craver and Darden, 2001], and Wright and Bechtel's "Mechanism" in this volume.) This work has yielded useful tools for dealing with what would, from the standpoint of traditional discussions of intertheoretic reduction, seem to be the more rarified distinctions between analytical levels in the sciences. Focusing on the details of particular mechanisms, they offer a bottom-up approach that suggests, if anything, delineating analytical levels on a case-by-case basis. (See too [Bickle, 2003, 116–117].) These accounts of mechanisms in science mostly leave it to philosophers of science interested in the

[8]Because Kim's discussion [1998, ch. 4] does not explore recent accounts of cross-scientific relations, explanatory pluralism, or mechanistic analysis in the philosophy of science or of ongoing research programs in the relevant sciences, it only sketches in the most abstract terms a similar sort of negative case concerning the promise of anti-reductive materialism.

more traditional questions of reductionism to worry about whether these instances will serve up patterns capable of supporting new generalizations about analytical levels at these finer resolutions.

These analyses of mechanisms exemplify the general morals of the explanatory pluralism that, according to the model of intertheoretic relations that has emerged across the previous four sections, prevails in nearly all cross-scientific settings. (See figure 5.) They offer, in effect, an explanatory pluralism writ *small*. They provide *multi-level* causal explanations that "... explain by showing *how* an event fits into a *causal* nexus" [Craver, 2001, 68]. It is the discovery and the delineation of the mechanism and its operations that reveals what researchers are willing to count as causal.

Bechtel [1986b] and Carl Craver [2001, 62–68] argue that analyses of mechanisms require inquiries that adopt at least three different perspectives. Minimally, the analysis of mechanisms will include and integrate studies of them and their operations as *isolated*, as *constituted*, and as *situated*. Craver prefers to distinguish these three perspectives from fixed levels of organization in nature: "these are three different perspectives on [an] . . . . item's activity in a hierarchically organized mechanism; they are not levels of nature" [2001, 66]. To stress the salience of the details of particular cases and to be (justifiably) wary about drawing any generalizations on the basis of particular cases do not, however, impugn their relevance to judgments about the finer grained analytical levels in science that have traditionally informed discussions of reduction. That mechanistic analyses concentrate on the local does not preclude their ability to illuminate reductionists' concerns about analytical levels.

To study a mechanism *as an isolated system* involves formulating a hypothesis about the system's borders, offering a rough and ready characterization of the mechanism's activities, and investigating the character, frequencies, locations, etc. of the inputs and outputs that cross those borders. Descriptions of mechanisms in isolation set constraints on approaches that examine their constituents.

The study of mechanisms *as constituted*, of course, is the most obvious point of contact with traditional discussions of microreduction, its concern with mereological relationships, and traditional concerns about levels of organization in nature. Whether the components in question are items treated in theories at a recognizably lower level usually depends on the scale of the breakdown. Functionally speaking, though, their study, regardless of the scale of the breakdown, exploits a more fine-grained perspective. According to both explanatory pluralism and Bechtel and Craver's analyses of mechanisms, reductive explanatory strategies are a fundamental part of the explanatory story in science, but they are not the entire story.

As noted earlier, complex systems can also be studied in isolation, but they rarely, if ever, operate in isolation. Examining a mechanism's performance in the larger environment of which it is a part provides explanatory insights of a different order. In context, the mechanism is now construed as part of a larger economy, which is studied at a functionally higher analytical level. Studies of mechanisms *as*

situated provide information about the roles that the mechanism as a whole plays in a larger dynamic system, about the sources of its inputs and the recipients of its outputs, and about other mechanisms that are capable of producing or influencing its inputs or outputs. Scrutinizing mechanisms in context commonly calls for appraisals of the functions of relationships and, at least from the biological level on up, of cooperative, competitive, and selective considerations especially.

Neither explanatory pluralism, writ large, nor these mechanistic analyses Polger cites leave any room for the dire or dismissive conclusions New Wave reductionists sometimes draw about intertheoretic relations in *interlevel settings*. All of the complex mechanisms and systems that populate the biological, psychological, cognitive, and socio-cultural sciences are, from the standpoint of explanation, greater than the sums of their parts and demand study as isolated, as constituted, and as situated. Such multi-level study not only offers a richer account of these particular mechanisms and their operations, it also enhances our understanding of all of the systems engaged in a mechanistic hierarchy. After all, these three perspectives are always *relative* to the particular mechanism that is the focal object of study. The strategies and pursuits at each level regularly yield findings that are mutually reinforcing [Craver, 2002]. Because they countenance the full range of possible cross-scientific relationships, these versions of explanatory pluralism do not confine themselves to reductive analyses only. Reductive explanation is a valuable contributor, but these approaches also embrace forms of non-reductive explanation as well [Polger, 2004, 205–209]. For the explanatory pluralist, all explanations are partial explanations; all explanations are from some perspective, and all explanations are motivated by and respond to specific problems.

10 HEURISTIC IDENTITY THEORY AND THE EXPLANATORY GAP OBJECTION

Discovering and explaining mechanisms proceeds piecemeal. Parallel research at multiple levels leads to increasingly developed views of mechanisms' organization and operations. Nothing intrinsic to analyses of mechanisms entails any particular ontological commitments, and Polger is correct that the models of these philosophers of science are "neutral about the nature of the entities that figure in mechanistic explanations" [2004, 209]. Nothing, however, follows about the implicit ontological commitments of scientists' specific mechanistic proposals. Progressively more integrated accounts of the structure and functioning of mechanisms yields the increasingly more articulated connections between models and theories at different analytical levels in science that animate philosophical naturalists' ontological calls.

Hypothesizing about cross-scientific identities serves as one of the principal heuristics for promoting such integrative research and for provoking discoveries at both of the explanatory levels involved. Polger [2004, 210] points out that "the identity theory in fact has more explanatory resources than functionalism because it makes use of both contextual and constitutive explanations." Such hypothetical

identities are common means for enabling scientists working at one analytical level to explore and exploit the conceptual, theoretical, methodological, and evidential resources available at another. The primary motives that drive the initial formulation of such hypotheses, whatever their eventual ontological consequences, concern their capacities to advance empirical research.

The logic behind their use looks to the converse of Leibniz's law. Instead of appealing to the identity of indiscernables, this strategy capitalizes on the indiscernability of identicals. What is known about an entity or process under one description should apply to it under its other descriptions. Scientists do not advocate hypothetical identities because the two characterizations *currently* mirror one another perfectly. On the contrary, it is just because the characterizations do not seem to mirror one another perfectly that the hypotheses are of interest. The theories at each level ascribe features to the entities and processes the interlevel, hypothetical identities connect. Scientists check the applicability of the feature lists at the alternative levels of analysis, using related research at each explanatory level to stimulate discovery at the other.

If unresolved conflicts persist, scientists pursue further empirical inquiry to ascertain either which characterization should prevail or in what directions each needs to evolve. *That* research yields more narrow hypotheses about the systems and patterns engaged. Within their respective levels, such speculations suggest new ways of organizing old theoretical commitments and familiar facts and point to new directions for empirical investigation. Finally, some of the new arrangements that result, almost inevitably, will produce some new cross-scientific conflicts, which will likely begin this cycle anew.

If, on the other hand, these hypotheses meet with much success at all, they will lead even more directly to the development and elaboration of more extended proposals about the connections between the two explanatory levels. Crucially, scientists accept or reject these hypotheses for the same reasons that they accept or reject any other hypotheses in science, viz., their abilities to stand up to empirical evidence, to stimulate new research, and to foster the integration of existing knowledge.

For example, in a relatively brief period in the middle of the twentieth century, research integrating behavioral and physiological techniques had defeated the hypothesized identification of the visual center with V1 exclusively. However, rather than undermining the strategy of hypothesizing identities, determining the functions of cells in V1 by David Hubel and Thorsten Wiesel [1962] led to more hypothetical identities that were even more precise about the visual and the neural processes identified and about the additional brain areas involved. Beginning with their landmark work and continuing for the next three decades, the ongoing interplay of behavioral and neurophysiological research led to repeated revisions of the hypotheses about the brain areas implicated and about the conception of the information processing performed in vision. This co-evolutionary process not only preserved a plurality of explanatory perspectives but resulted in a refinement of both psychological and neural models. For over a century now, exploring the

empirical merits of the hypothesized identity of the visual center and the occipital lobe has generated both more and more detailed hypotheses about the activities in the brain with which various aspects of visual processing and experience should be identified. Moreover, these findings about the various brain areas involved in visual processing played a key role in inspiring new, ambitious, higher-level hypotheses at the neuropsychological and psychological levels about visual processing and the organization of the human cognitive system overall. Probably the most famous is Ungerleider and Mishkin's [1982] identification of two processing streams for visual information in the brain.

Bechtel and McCauley appropriate these lessons from the philosophy of science and from their instantiations at the interface of psychology and neuroscience, in particular, to formulate both a new version of the psycho-physical identity theory and replies to the multiple realizability objection (scouted at the end of section 8) and to the explanatory gap objection ([Bechtel and McCauley, 1999]; [McCauley and Bechtel, 2001]). According to their Heuristic Identity Theory (HIT) psychoneural identities are not the conclusions of scientific research but the hypothetical premises. The preceding discussions of explanatory pluralism and mechanistic analysis show why hypothetical identities of psychological and neural processes generate both new hypotheses and new avenues of research that serve to direct those hypotheses' development and elaboration. The differences between theories at these two levels encourage scientists to consider adjustments to their conceptions of the pertinent processes and structures in a reciprocal process of mutual fine-tuning. By way of illustration, Bechtel and McCauley review the history of proposals about the locus of visual processing in the brain from the late nineteenth through the late twentieth centuries (briefly touched upon above).

The way that HIT construes psycho-physical identities suggests a framework for responding to the objection to the reduction of psychology (and to the identity theory) that appeals to an explanatory gap concerning consciousness. On HIT's account of things, finally, an explanatory gap at the interface of psychology and neuroscience, whatever its basis, is simply an instance of a failure of reductive integration between two sciences operating at adjacent analytical levels. The gap will be closed the same way that other cross-scientific gaps have been closed in the history of science, viz., through the co-evolutionary integration of the sciences in the course of on-going cross-scientific research.

Kim nicely summarizes the general argument informing the objection[9]: "*it is*

[9]David Chalmers' version of the argument [1996, 115 goes as follows:

> Neurobiological approaches to consciousness ... can ... tell us something about the brain processes that are *correlated* with consciousness. But none of these accounts explains the correlation: we are not told why brain processes should give rise to experience at all. From the point of view of neuroscience, the correlation is simply a brute fact.
>
> ... Because these theories gain their purchase by *assuming* a link ... it is clear that they do nothing to explain that link.

For an extended reply to this specific formulation of the explanatory gap objection, see [McCauley and Bechtel, 2001].

the explanation of ... bridge laws, an explanation of why there are just these mind-body correlations, that is at the heart of the demand for an explanation of mentality ... it is evident that the Nagel reduction of psychology is like taking mind-body supervenience as an unexplained brute fact" [1998, 96]. Within the objection lurk two challenges: (1) the identity theory and reductive materialism need to explain the identities they propose and (2) any evidence that can be cited in support of such an explanation is also perfectly consistent with affirming no more than psychoneural correlations (and, thus, anti-reductionists fault both positions for their metaphysical presumption). The specific complaint about an explanatory gap concerning consciousness maintains that physicalist accounts provide no explanation, in particular, of how something psychic could *just be* something physical.

The second challenge amounts to arguing that, from the standpoint of the logic of confirmation, claims about the identity of two things are indistinguishable from claims about their correlation. As Kim [1966, 227] has put the objection: "... the factual content of the identity statement is exhausted by the corresponding correlation statement ... There is no conceivable observation that would confirm or refute the identity but not the associated correlation." If the whole philosophical story about proposing psycho-physical identities were one about their confirmation, then this perfectly uncontroversial logical point would carry the day. HIT, however, maintains that cross-scientific hypothetical identities between psychological and neural processes are part of a multi-level scientific investigation of human mentality that involves much more than connections between theories' ontologies (or their confirmation). Claims about correlations and claims about interlevel identities are different conceptual animals that thrive in different theoretical habitats. Not only does this part of the objection overlook the fundamental contribution hypothetical identities make to scientific *discovery*, it does not even get the role of these identities right in the *justification* of scientific theories. Unlike merely noting correlations, advancing hypothetical identities occasions explanatory connections that *demand* empirical exploration. Cross-scientific identities make evidence available from other explanatory levels, and, as noted above, they disclose avenues of research for generating new evidence as well. Their critical contribution resides in their abilities to provoke and refine theories at both of the levels engaged. *Their success at this task is their vindication — not* the accumulation of some sort of evidence that would rule the corresponding correlation claim out of court. The whole point of the correlation objection is precisely that such evidence cannot exist!

HIT simply denies the assumption underlying the first challenge. Identities are not the sorts of things that ever need explanation. What matters about hypothetical cross-scientific identities is not how they should be explained (they can't be) but what they explain, how they suggest (and contribute to) other, empirically successful, explanatory hypotheses, and how they create opportunities for scientists who work at one explanatory level to enlist methods and evidence from alternative levels of explanation. *That* is why "we are not told why brain processes should give

rise to experience ..." [Chalmers, 1996, 115]. Scientists show why some mechanism *constitutes* some phenomenon by exploring the empirical success of the wide range of predictions and explanatory connections that assumption generates. It is that empirical success that corroborates the constitutive hypothesis and tentatively justifies its assumption [Churchland and Churchland, 1998, 120–122]. But, of course, the tentativeness here is nothing special. It is the same tentativeness about justification that accompanies every empirical claim in science.

HIT underscores the fact that evaluations of proposed identities do not turn on confirming them directly. What, after all, could *that* possibly be [McCauley, 1981]? In empirical matters the evidence for an identity claim arises *indirectly* — primarily on the basis of the emerging empirical support for the explanatory hypotheses it informs. For example, if normal activities in V4 are identical with the processing of information about wave length, then serious abnormalities of particular types in the structure and functioning of V4 should yield abnormalities of particular types in subjects' color experiences. The point is that this hypothetical identity is an empirical conjecture that researchers can use both psychological and neuroscientific evidence not only to assess but to refine. Obtaining indirect corroborating evidence for identifying some neural process with some psychological function along such lines no more finalizes that identity than it would any other hypothesis in science. Nor does it establish that the function under scrutiny is either the sole or even the primary function these neural processes carry out. (So, in fact, whether V4 is even primarily concerned with the processing of color is a point of some controversy among researchers.) Moreover, all research of this sort is limited by scientists' abilities both to conceive of what stimuli might provoke responses in a neural area and to test those conceptions. Still, the more hypotheses of this sort the identity informs and the more successful those hypotheses prove, the more likely the identity will come to serve as an assumption the sciences lean upon rather than a bare conjecture in search of support [Van Gulick, 1997]. Such identity claims are, of course, no less conjectures still. They are, however, no longer simply *bare* conjectures. Nor, manifestly, are the identities that HIT surveys "brute," contrary both to other versions of the psycho-physical identity theory and to virtually all of these anti-reductionist critics.

Those critics miss both the sorts of considerations that motivate hypothetical identities in science and their fundamental contribution to the development of scientific explanations. An emphasis on the multi-level character of scientific research in the sciences of the mind/brain does not bar the explanatory pluralist from embracing type-identities of suitable granularity between mental processes and brain processes. The explanatory and predictive progress that such hypothetical identities promote is the best reason available for acknowledging such cross-scientific connections.

BIBLIOGRAPHY

[Bartlett, 1923] F. C. Bartlett. *Remembering*. Cambridge: Cambridge University Press, 1923.

[Bechtel, 1986a] W. Bechtel, ed. *Integrating Scientific Disciplines*. The Hague: Martinus Nijhoff, 1986.

[Bechtel, 1986b] W. Bechtel. Teleological Functional Analyses and the Hierarchical Organization of Nature, *Teleology and Natural Science*. N. Rescher (ed.). Landham, MD: University Press of America, 26-48, 1986.

[Bechtel and McCauley, 1999] W. Bechtel and R. N. McCauley. Heuristic Identity Theory (or Back to the Future): The Mind-Body Problem Against the Background of Research Strategies in Cognitive Neuroscience, *Proceedings of the Twenty-First Meeting of the Cognitive Science Society*. M. Hahn and S. C. Stones (eds.). Mahway, New Jersey: Lawrence Erlbaum Associates, 67-72, 1999.

[Bechtel and Mundale, 1999] W. Bechtel and J. Mundale. Multiple Realizability Revisited: Linking Cognitive and Neural States, *Philosophy of Science* 66, 175-207, 1999.

[Bechtel and Richardson, 1993] W. Bechtel and R. C. Richardson. *Discovering Complexity: Decomposition and Localization as Strategies in Scientific Research*. Princeton: Princeton University Press, 1993.

[Bickle, 1998] J. Bickle. *Psychoneural Reduction: The New Wave*. Cambridge, MA: MIT Press, 1998.

[Bickle, 2003] J. Bickle. *Philosophy and Neuroscience: A Ruthlessly Reductive Account*. Dordrecht: Kluwer Academic Publishers, 2003.

[Block, 1997] N. Block. AAnti-Reductionism Slaps Back,@ *Philosophical Perspectives* 11, 107-132, 1997.

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

[Broca, 1861] P. Broca. ARemarque sur le Siêge de la Faculté Suivies d=une Observation "Aphémie", *Bulletins de la Société Anatomique de Paris* 6, 343-357, 1861.

[Brodman, 1909/1994] K. Brodmann. *Vergleichende Lokalisationslehre der Grosshirnrinde*. L. J. Garvey, trans. Leipzig: J. A. Barth, 1909/1994.

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

[Chalmers, 1995] D. Chalmers. Facing Up to the Problem of Consciousness, *Journal of Consciousness Studies* 2, 200-219, 1995.

[Chalmers, 1996] D. Chalmers. *The Conscious Mind: In Search of a Fundamental Theory*. New York: Oxford University Press, 1996.

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

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

[Churchland and Churchland, 1996] P. M. Churchland and P. S. Churchland. Replies from the Churchlands, *The Churchlands and Their Critics*. R. N. McCauley (ed.). Oxford: Blackwell Publishers, pp. 217-310, 1996.

[Churchland and Churchlan, 1998] P. M. Churchland and P. S. Churchland. Intertheoretic Reduction: A Neuroscientist's Field Guide, *On the Contrary*. Cambridge, MA: MIT Press, 65-79, 1998.

[Churchland, 1983] P. S. Churchland. Consciousness: The Transmutation of a Concept, *Pacific Philosophical Quarterly* 64, 80-93, 1983.

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

[Churchland, 1988] P. S. Churchland. Reduction and the Neurological Basis of Consciousness, *Consciousness in Contemporary Science*. A. J. Marcel, and E. Bisiach (eds.). Oxford University Press, Oxford, 273-304, 1988.

[Churchland, 2002] P. S. Churchland. *Brain-Wise: Studies in Neurophilosophy*. Cambridge: MIT Press, 2002.

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

[Craver, 2001] C. F. Craver. Role Functions, Mechanisms, and Hierarchy, *Philosophy of Science* 68, 53-74, 2001.

[Craver, 2002] C. F. Craver. AInterlevel Experiments and Multilevel Mechanisms in the Neuroscience of Memory,@ *Philosophy of Science*, supplement to volume 69, S83-S97, 2002.

[Craver and Darden, 2001] C. F. Craver and L. Darden. Discovering Mechanisms in Neuroscience: The Case of Spatial Memory, *Theory and Method in Neuroscience*. P. Machamer, R. Grush, and P. McLaughlin (eds.). Pittsburgh, PA: University of Pittsburgh Press, 2001.
[Cummins, 2000] R. Cummins. 'How Does It Work?' versus 'What Are the Laws?': Two Conceptions of Psychological Explanation, *Explanation and Cognition*. F. Keil and R. Wilson (eds.). Cambridge: MIT Press, 117-144, 2000.
[Cunningham, 2001a] B. Cunningham. The Reemergence of 'Emergence', *Philosophy of Science*, 68 (Supplement-Proceedings), S62-S75, 2001.
[Cunningham, 2001b] B. Cunningham. Capturing Qualia: Higher-Order Concepts and Connectionism, *Philosophical Psychology* 14, 29-41, 2001.
[Darden and Maull, 1977] L. Darden and N. Maull. Interfield Theories, *Philosophy of Science* 44, 43-64, 1977.
[Davidson, 1970] D. Davidson. Mental Events, *Experience and Theory*. L. Foster and J. Swanson (eds.) Amherst, MA: University of Massachusetts Press, 79-101, 1970.
[Ebbinghaus, 1964/1885] H. Ebbinghaus. *Memory: A Contribution to Experimental Psychology*. New York: Dover, 1964/1885.
[Farah *et al.*, 1998] M. J. Farah, K. D. Wilson, H. M. Drain and J. R. Tanaka. What is 'Special' about Face Perception? *Psychological Review* 105, 482-498, 1998.
[Ferier, 1876] D. Ferrier. *The Functions of the Brain*. London: Smith, Elder, and Company, 1876.
[Feyerabend, 1962] P. K. Feyerabend. Explanation, Reduction, and Empiricism, *Minnesota Studies in the Philosophy of Science, Volume III*. H. Feigl and G. Maxwell (eds.). Minneapolis: University of Minnesota Press. 28-97, 1962.
[Flanagan, 1992] O. Flanagan. *Consciousness Reconsidered*. Cambridge: The MIT Press, 1992.
[Fodor, 1974] J. A. Fodor. Special Sciences (or: The Disunity of Science as a Working Hypothesis), *Synthese* 28, 97-115, 1974.
[Fodor, 1975] J. A. Fodor. *The Language of Thought*. New York: Thomas Y. Crowell Company, 1975.
[Fodor, 1981] J. A. Fodor. The Mind-Body Problem, *Scientific American* 244, 114-123, 1981.
[Fodor, 1983] J. A. Fodor. *The Modularity of Mind*. Cambridge: The MIT Press, 1983.
[Fodor and Pylyshyn, 1988] J. A. Fodor and Z. W. Pylyshyn. Connectionism and Cognitive Architecture: A Critical Analysis. *Cognition* 28, 3-71, 1988.
[Glenberg, 1976] A. M. Glenberg. Monotonic and Nonmonotonic Lag Effects in Paired-Associate and Recognition Memory Paradigms, *Journal of Verbal Learning and Verbal Behavior* 15, 1-16, 1976.
[Hardcastle, 1996] V. G. Hardcastle. *How to Build a Theory in Cognitive Science*. Albany: SUNY Press, 1996.
[Henschen, 1893] S. Henschen. On the Visual Path and Centre, *Brain* 16, 170-180, 1893.
[Hirst and Gazzaniga, 1988] W. Hirst and M. S. Gazzaniga. Present and Future of Memory Research and Its Applications, *Perspectives in Memory Research*. M. S. Gazzaniga (ed.). Cambridge: The MIT Press, 1988.
[Hooker, 1981] C. Hooker. Towards a General Theory of Reduction, *Dialogue* 20: 38-59, 201-36, 496-529, 1981.
[Hubel and Wiesel, 1962] D. H. Hubel and T. N. Wiesel. Receptive Fields, Binocular Interaction and Functional Architecture in the Cat's Visual Cortex, *Journal of Physiology* (London) 160, 106-154, 1962.
[Jackson, 1982] F. Jackson. Epiphenomenal Qualia, *Philosophical Quarterly* 32, 127-36, 1982.
[Jackson, 1986] F. Jackson. What Mary Didn't Know, *Journal of Philosophy* 83, 291-95, 1986.
[Jacoby, 1978] L. L. Jacoby. On Interpreting the Effects of Repetition: Solving a Problem Versus Remembering a Solution, *Journal of Verbal Learning and Verbal Behavior* 17, 649-667, 1978.
[Kanwisher *et al.*, 1997] N. Kanwisher, J. McDermott, and M. M. Chun. AThe Fusiform Face Area: A Module in Human Extrastriate Cortex Specialized for Face Perception,@ *Journal of Neuroscience* 17, 4302-4311, 1997.
[Kemeny and Oppenheim, 1956] J. Kemeny and P. Oppenheim. On Reduction, *Philosophical Studies* 7, 6-19, 1956.
[Kim, 1966] J. Kim. On the Psycho-Physical Identity Theory, *Materialism and the Mind-Body Problem*. D. Rosenthal (ed.). Englewood Cliffs, New Jersey: Prentice-Hall, 80-95, 1966.
[Kim, 1972] J. Kim. Phenomenal Properties, Psychophysical Laws, and the Identity Theory, *Monist* 56, 177-192, 1972.

[Kim, 1989] J. Kim. The Myth of Nonreductive Materialism, *Proceedings of the American Philosophical Association* 63, 3, pp. 31-47, 1989.
[Kim, 1998] J. Kim. *Mind in a Physical World: An Essay on the Mind-Body Problem and Mental Causation*. Cambridge: MIT Press, 1998.
[Kuhn, 1970] T. Kuhn. *The Structure of Scientific Revolutions* (second edition). Chicago: University of Chicago Press, 1970.
[Levine, 1983] J. Levine. Materialism and Qualia: The Explanatory Gap, *Pacific Philosophical Quarterly* 64, 354-61, 1983.
[Levine, 1977] J. Levine. On Leaving Out What It's Like, *The Nature of Consciousness: Philosophical Debates*. N. Block, O. Flanagan, and G. Güzeldere (eds.). Cambridge: MIT Press, 1977.
[Looren de Jong, 1997] H. Looren de Jong. Levels: Reduction and Elimination in Cognitive Neuroscience, *Problems of Theoretical Psychology*. C. W. Tolman, F. Cherry, R. van Hezewijk, I. Lubek (eds.). New York: Captus Press, pp. 165-172, 1997.
[Machamer *et al.*, 2000] P. Machamer, L. Darden, and C. F. Craver. Thinking about Mechanisms, *Philosophy of Science* 67, 1-25, 2000.
[Maull, 1977] N. Maull. Unifying Science Without Reduction, *Studies in History and Philosophy of Science* 8, 143-62, 1977.
[McCauley, 1981] R. N. McCauley. Hypothetical Identities and Ontological Economizing: Comments on Causey's Program for the Unity of Science, *Philosophy of Science* 48, 218-27, 1981.
[McCauley, 1986] R. N. McCauley. Intertheoretic Relations and the Future of Psychology, *Philosophy of Science* 53, 179-99, 1986.
[McCauley, 1996] R. N. McCauley. Explanatory Pluralism and the Coevolution of Theories in Science, *The Churchlands and Their Critics*. R. N. McCauley (ed.). Oxford: Blackwell Publishers, 17-47, 1996.
[McCauley and Bechtel, 2001] R. N. McCauley. and W. Bechtel. Explanatory Pluralism and The Heuristic Identity Theory, *Theory and Psychology* 11, pp. 738-761, 2001.
[McGaugh, 2000] J. McGaugh. Memory: A Century of Consolidation, *Science* 287, 248-251, 2000.
[McGinn, 1991] C. McGinn. *The Problem of Consciousness*. Oxford: Blackwell Publishers, 1991.
[Mishkin *et al.*, 1983] M. Mishkin, L. G. Ungerleider, and K. A. Macko. Object Vision and Spatial Vision: Two Cortical Pathways, *Trends in Neurosciences* 6, 414-417, 1983.
[Mundale, 2001] J. Mundale. Neuroanatomical Foundations of Cognition: Connecting the Neuronal Level with the Study of Higher Brain Areas, *Philosophy and the Neurosciences: A Reader*. W. Bechtel, P. Mandik, J. Mundale, and R. S. Stufflebeam (eds.). Oxford: Blackwell Publishers, 37-54, 2001.
[Mundale and Bechtel, 1996] J. Mundale and W. Bechtel. Integrating Neuroscience, Psychology, and Evolutionary Biology through a Teleological Conception of Function, *Minds and Machines* 6, 481-505, 1996.
[Nagel, 1961/1979] E. Nagel. *The Structure of Science*. New York: Harcourt, Brace and World / Indianapolis: Hackett, 1961/1973.
[Neisser, 1967] U. Neisser. *Cognitive Psychology*. New York: Appleton-Century-Crofts, 1967.
[Oppenheim and Putnam, 1958] P. Oppenheim and H. Putnam. Unity of Science as a Working Hypothesis, *Minnesota Studies in the Philosophy of Science-Volume II*, 1958.
[Polger, 2004] T. Polger. *Natural Minds*. Cambridge: MIT Press, 2004.
[Putnam, 1967/1975] H. Putnam. Psychological Predicates, *Art, Mind, and Religion*. W. Capitan and D. Merrill (eds.). Pittsburgh: University of Pittsburgh Press / reprinted in *Mind, Language and Reality: Philosophical Papers* (volume 2). New York: Cambridge University Press, 1967/1975.
[Richardson, 1979] R. Richardson. Functionalism and Reductionism, *Philosophy of Science* 46, 533-558, 1979.
[Richardson, 1982] R. Richardson. How *Not* to Reduce a Functional Psychology, *Philosophy of Science* 49, 125-37, 1982
[Schaffner, 1967] K. Schaffner. Approaches to Reduction, *Philosophy of Science* 34, 137-47, 1967.
[Schaffner, 1992] K. Schaffner. Philosophy of Medicine, *Introduction to the Philosophy of Science*. M. Salmon, J. Earman, C. Glymour, J. Lennox, P. Machamer, J. McGuire, J. Norton, W. Salmon, and K. Schaffner. Englewood Cliffs, New Jersey: Prentice Hall, 310-344, 1992.

[Schouten and Looren de Jong, 1999] M. K. D. Schouten and H. Looren de Jong. Reductionism, Elimination, and Levels: The Case of the LTP-Learning Link, *Philosophical Psychology* 12, 237-262, 1999.
[Sejnowski and Rosenberg, 1987] T. J. Sejnowski and C. R. Rosenberg. Parallel Networks that Learn to Pronounce English Text, *Complex Systems* 1, 145-168, 1987.
[Sejnowski and Rosenberg, 1988] T. J. Sejnowsk and C. R. Rosenberg. Learning and Representation in Connectionist Models, *Perspectives in Memory Research*. M. Gazzaniga (ed.). Cambridge: The MIT Press. 135-78, 1988.
[Simon, 1969] H. Simon. *The Sciences of the Artificial*. Cambridge: MIT Press, 1969.
[Thagard, 1992] P. Thagard. *Conceptual Revolutions*. Princeton: Princeton University Press, 1992.
[Ungerleider and Mishkin, 1982] L. G. Ungerleider and M. Mishkin. Two cortical visual systems. D. J. Ingle, M. A. Goodale, and J. W. Mansfield (eds.), *Analysis of Visual Behavior*. Cambridge, MA: MIT Press, 549586, 1982.
[van Essen and Gallant, 1994] D. C. van Essen and J. L. Gallant. Neural Mechanisms of Form and Motion Processing in the Primate Visual System, *Neuron* 13, 1-10, 1994.
[Van Gulick, 1997] R. Van Gulick. Understanding the Phenomenal Mind: Are We All Just Armadillos? *The Nature of Consciousness: Philosophical Debates*. N. Block, O. Flanagan, and G. G(zeldere (eds.). Cambridge: MIT Press, 1997.
[Wimsatt, 1974] W. Wimsatt. Complexity and Organization, *PSA-1972*. K. Schaffner and R. Cohen (eds.). Dordrecht: Reidel, 67-82, 1974.
[Wimsatt, 1976] W. Wimsatt. Reductionism, Levels of Organization, and the Mind-Body Problem, *Consciousness and the Brain*. G. Globus, G. Maxwell, and I. Savodnik (eds.). New York: Plenum Press. 205-67, 1976.
[Wimsatt, 1978] W. Wimsatt. Reduction and Reductionism, *Current Problems in Philosophy of Science*. P. Asquith and H. Kyburg (eds.). East Lansing: Philosophy of Science Association. 1-26, 1978.
[Wimsatt, 1986] W. Wimsatt. Forms of Aggregativity, *Human Nature and Natural Knowledge*. M. Wedin (ed.). Dordrecht: Reidel, 259-293, 1986.
[Wimsatt, 1997] W. Wimsatt. Aggregativity: Reductive Heuristics for Finding Emergence, *Philosophy of Science* 64, S372-S384, 1997.

PERCEPTION PREATTENTIVE AND PHENOMENAL

Austen Clark

The conundrums of phenomenal character and consciousness have often motivated philosophers to study perception; and a particular area of study that will prove worthy of their attention is research into what are called "early" or "preattentive" perceptual processes. These are, roughly, processes that start at the transducers and end where selective attention has access to the results, and can select some favored few for further processing. These early processes are the ones most likely to be called "sensory"; they are at any rate simpler, and they make their appearance earlier than, the more sophisticated states that underlie perceptual judgments. If non-conceptual representation is employed anywhere in the system, it would be employed here. Animals that cannot muster the words for a perceptual judgement can nevertheless sense things. These simple sensory states are dear to the heart of those interested in phenomenal character, and I hope they find their interests piqued by preattentive phenomena.

To fit constraints of time and space this paper must set aside consideration of relations between perception and knowledge, and between perception and action, even though there is enormously interesting work underway on both fronts. We will mine single-mindedly the vein that leads into phenomenology. To narrow the topic even further, I will confine the discussion to early vision. It is still an enormous field, and it provides more than enough materiel with which to examine some recent discussions of phenomenal consciousness.

The reader should be forewarned: the architecture of early vision is surprising, bizarre, *weird*. I will describe some of the surprising and weird features of that architecture, and then consider how folk concepts of appearance and awareness might be applied to it. The results, like the architecture itself, are somewhat bizarre. In particular, the preattentive architecture challenges the idea – it puts enormous stress on the common sense notion – that "phenomenal character" is and must be coeval with awareness. Instead, the two split apart, with "phenomenal character" showing up first, in places as yet unoccupied by awareness. After reviewing some of the evidence, I shall argue that the most reasonable conclusion is that some states of preattentive sensing are states in which one is being appeared-to, even though one is entirely unaware of what appears, of any aspect of its appearance, or of being in the state of being appeared-to. So "phenomenal character" and consciousness fly apart; their association in "phenomenal consciousness" is a merely contingent conjunction of two distinct things. This puts some stress on our ordinary notions; one burden of the argument is to consider the least costly methods for stress relief.

Handbook of the Philosophy of Science. Philosophy of Psychology and Cognitive Science
Volume editor: Paul Thagard
General editors: Dov M. Gabbay, Paul Thagard and John Woods