Robert Van Gulick

Reduction, Emergence and Other Recent Options on the Mind/Body Problem

A Philosophic Overview

I: Introduction

Though most contemporary philosophers and scientists accept a physicalist view of mind, the recent surge of interest in the problem of consciousness has put the mind/body problem back into play. The physicalists' lack of success in dispelling the air of residual mystery that surrounds the question of how consciousness might be physically explained has led to a proliferation of options. Some offer alternative formulations of physicalism, but others forgo physicalism in favour of views that are more dualistic or that bring in mentalistic features at the ground-floor level of reality as in pan-proto-psychism.

The situation might be viewed as a case of what the philosopher of science Thomas Kuhn (1962) called *extraordinary science*, i.e., a period of ferment and theoretical experimentation in which the practices and concepts of so-called 'normal science' are loosened in ever more radical ways in response to a persistent anomaly, a problem that resists solution within the limits of the prevailing *normal science* paradigm. Some such episodes usher in full-blown scientific revolutions that sweep away and supplant the prior paradigm; others get resolved more conservatively as ways are found to resolve the anomaly without wholesale abandonment of prior commitments. In those Kuhnian terms, physicalism (particularly the sort of functionalistic nonreductive physicalism that has become the mainstream view among philosophers in recent decades) plays the role of normal science, and consciousness (especially the so-called 'hard problem' of explaining how phenomenal consciousness might be just a physical aspect of reality) provides the anomaly that generates the push toward extraordinary theorizing (Chalmers, 1996). How the current psycho-physical crisis will be resolved as yet remains unclear; revolutions may or may not be needed.

Journal of Consciousness Studies, **8**, No. 9–10, 2001, pp. 1–34

My aim here is to give an overview of the recent philosophic discussion to serve as a map in locating issues and options. I will not offer a comprehensive survey of the debate or mark every important variant to be found in the recent literature. I will mark the principal features of the philosophic landscape that one might use as general orientation points in navigating the terrain.

I will focus in particular on three central and interrelated ideas: those of emergence, reduction, and nonreductive physicalism. The third of these, which has emerged as more or less the majority view among current philosophers of mind, combines a pluralist view about the diversity of what needs to be explained by science with an underlying metaphysical commitment to the physical as the ultimate basis of all that is real. The view has been challenged from both left and right, on one side from dualists (Chalmers, 1996) and on the other from hard core reductive materialists (Kim, 1989). Despite their differences, those critics agree in finding nonreductive physicalism an unacceptable and perhaps even incoherent position. They agree as well in treating reducibility as the essential criterion for physicality; they differ only about whether the criterion can be met. Reductive physicalists argue that it can, and dualists deny it.

The terms 'reduction', 'nonreductive' and 'emergence' get used in a bewildering variety of ways in the mind–body literature, none of which is uniquely privileged or standard. Thus clarity about one's intended meaning is crucial to avoid confusion and merely verbal disagreements. Thus, much of my mapping will be devoted to sorting out the main versions of reduction and emergence before turning to assess their interrelations and plausibility. My intent is to act largely as a guide and not an advocate. Though I am sure my biases will sometimes affect how I describe the issues, my goal is to lay out the logical geography in a more-or-less neutral way.

II: The Varieties of Reduction

The basic idea of reduction is conveyed by the 'nothing more than . . .' slogan. If Xs reduce to Ys, then we would seem to be justified in saying or believing things such as 'Xs are nothing other (or more) than Ys', 'Xs are just special sorts, combinations or complexes of Ys', or 'Xs are nothing over and above Ys'. However, once one moves beyond slogans, the notion of reduction is ambiguous along two principal dimensions: the types of items that are reductively linked and the nature of the link involved. Thus, to define a specific notion of reduction, we need to answer two questions:

- Question of the relata: Reduction is a relation, but *what types of things* does it link?
- Question of the link: *In what way(s)* must the items be linked to count as a reduction?

Let us first consider the question of the relata. Between *what types of things* might the reduction relation hold? The notion gets interpreted in two distinct ways that involve very different sorts of relata. It can be viewed either as

- a relation between real-world items — objects, events, or properties — which we might term *Ontological Reduction* (ONT-Reduction).

or as

- a relation between representational items — theories, concepts or models — which we can call *Representational Reduction* (REP-Reduction).

There are obviously important connections between the two families. But they involve distinct types of relata, and one must not conflate them as too often happens. Speakers in the reduction debate often talk past one another by failing to distinguish ontological from representational notions, especially in interdisciplinary settings that combine scientists and philosophers.

The distinction is crucial as well for locating nonreductive physicalism in the logical space of options. It typically combines a denial of some form(s) of representational reduction with the acceptance of some type(s) of ontological reduction supposedly adequate to secure its physicalist credentials. It claims to coherently conjoin representational nonreduction with an ontological link robust enough to meet the demands of physicalism. It's critics deny that can be done, but the claim at least locates the view in logical space.

The diagram in figure 1 shows the first step in our taxonomy by subdividing the types of reduction into two families based on their answer to our first diagnostic question.

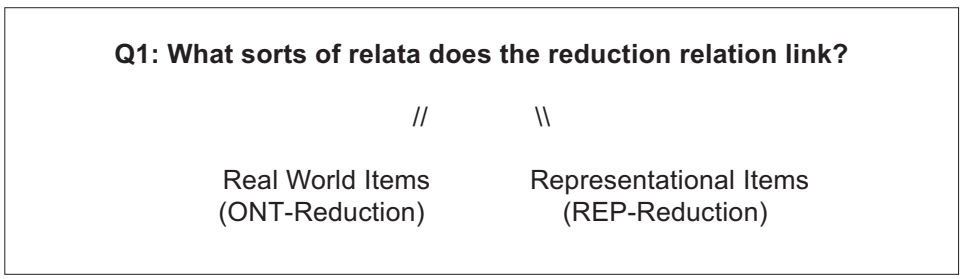

Figure 1

Each family further subdivides based on the specific types of relata involved. Thus, ontological reductions might involve relations between things of various kinds:

- objects in two domains (e.g., minds and brains, or pains and neuron firings),
- properties (e.g., feeling pain and having neural activation of type N_p, or wanting a cup of coffee and being in a neurofunctional state of type Nf_d),
- events (e.g., Bill's having a red visual experience and Bill's brain being globally active in a way that includes neural activities of type N_{rve} in his visual cortex as part of its global focus),
- processes (e.g., my recalling of the cellist's performance and a sequence of reciprocal neural interactions RNI_c between multiple limbic and cortical areas).

REP-reduction similarly divides into more specific subtypes based on the particular relata involved, which might include any of the following:

- concepts (e.g., links between our first-person concept of phenomenal red and concepts from neuroscience),
- theories (e.g., links between theories of conscious experience and theories of global brain function),
- models (e.g., links between models of consciousness and models of reciprocal brain activity),
- representational frameworks (e.g., links between the phenomenal first-person descriptive/explanatory framework and third-person neuroscience frameworks).

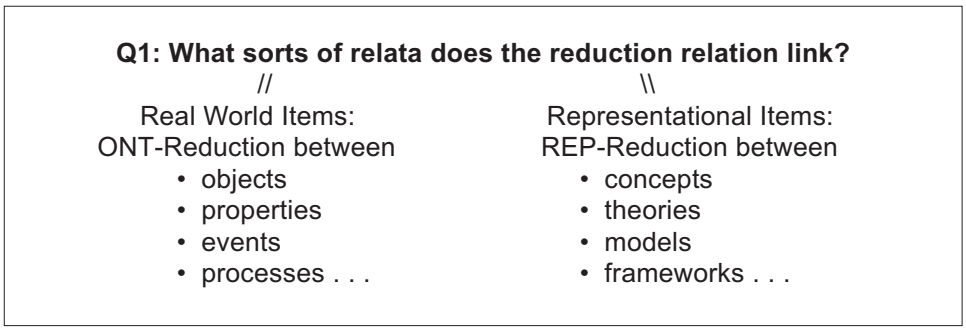

Figure 2

Let us move on to our second diagnostic probe, the question of the link: *In what way(s)* must the items be *linked* to count as a case of reduction? Again, there are a variety of answers on both the ontological and the representational side. With regard to ONT-reduction, the question becomes:

> Question of the *ontological link:* How must things be related for one to ontologically reduce to the other?

At least five major answers have been championed in the literature:

- elimination
- identity
- composition
- supervenience
- realization

The relative merits and faults of the competing proposals have been extensively and intricately debated, but for present purposes it should suffice to say a brief bit about each and give a general sense of the range of options.

Elimination. One of the three forms of reduction listed by Kemeny and Oppenheim in their classic paper on reduction (1956) was replacement, i.e., cases in which we come to recognize that what we thought were Xs are really just Ys. For example, we've come to see that what had been thought of as demon possession is just a form of psychosis, perhaps a type of schizophrenia with auditory hallucinations. In such a case we might say that demon possession has been reduced to a mental illness; schizophrenia replaces demon possession in our inventory of

the world. Eliminativists believe that a similar fate awaits many of the commonly alleged denizens of the mental domain; qualia, beliefs, intentionality and even consciousness have all been slated for eventual replacement by supposedly more mature scientific alternatives (Rorty, 1970; Churchland, 1981; Dennett, 1988; Wilkes, 1988; 1995). It is difficult to prove or refute predictions about the future course of inquiry and how it will affect our future conceptual repertoire and beliefs about what is and isn't real. Nonetheless, the eliminativist shoulders a nontrivial burden to motivate his claim that so much of what we take to be real in the mental domain might go the way of demons or black bile. Our sense of what is mentally real seems too intimate and too useful to turn out to be so badly mistaken as to justify an eliminativist judgment. The debate in the literature is extensive, but in this context it is enough just to note the controversial status of the issue.

Some contemporary readers may find it odd to describe elimination as a form of reduction since according to the eliminativist there is nothing really there to be reduced. If there are no beliefs, how can they be reduced to anything? Nonetheless, I believe it is appropriate to keep it on our list. As noted above, reduction by replacement, which is what elimination basically involves, was listed early on by Kemeny and Oppenheim in their seminal work (1956) as one of the main types of reduction. Moreover, the eliminativist view is one way to unpack the basic reductive notion of 'nothing but . . .'; demon possession (or what we thought of as demon possession) turns out to be nothing but organic brain disorder.

Identity. Identity falls at the opposite extreme from elimination. It involves cases in which we continue to accept the existence of Xs but come to see that they are identical with Ys (or with special sorts of Ys). Xs reduce to Ys in the strictest sense of being identical with Ys. This most often happens when a later Y-theory reveals the true nature of Xs to us. We have come to see that heat is just kinetic molecular energy, that lightning is just an atmospheric discharge of static electricity and that genes are just functionally active DNA sequences. However, the 'just are . . .' locution in such cases does not lead us to eliminate or deny the existence of the prior items; we do not deny the reality of lightning, heat or genes. Rather, we see that two distinct reference routes converge on the same item. In the logician Frege's famous example of the evening star and the morning star, both turned out to be the planet Venus. (Frege, 1892) So, too, identity theorists claim mental states, events and properties will turn out to be identical with neuroscientifically discovered items. Those mental states are every bit as real as heat and genes, but their true nature is yet to be discovered by a more mature mind/brain theory.

Contemporary physicalism first developed as an identity theory in the 1950s and 60s (Place, 1956; Smart, 1959), specifically as what has come to be called the type–type central state identity theory, which held that types of mental states (e.g., being a stabbing pain) were identical with types of states within the central nervous system (e.g., c-fibre firings or, more realistically, certain patterns of firing in anterior cingulate, somato-sensory cortex and interconnected limbic areas). The theory fell out of favour quickly for a variety of reasons, such as the multiple realizability objection which appeals to the fact that one and the same type of mental state might be realized by neurally quite dissimilar structures in different

creatures, in different humans, or even in the same person at different times (Putnam, 1972; Kim, 1993b). Thus, by the late 1960s, most physicalists had moved away from the type–type identity theory in favour of some form of functionalism that treats mental states and properties as higher-order features defined by their higher-level roles but nonetheless realized solely by their neural substrates, much as the higher-level program states of a computer are realized by the underlying states of its hardware.

The situation is actually a bit more complicated since some functionalists (Lewis, Armstrong, . . .) identify the mental state-type not with the higher-order or role property but with the specific structural property that plays that role in a given species or population. Though they pick out the property by the role it typically plays, they set the identity conditions in structural terms. In that respect they are more like classic type–type identity theorists than like other functionalists who set the identity conditions in terms of the role itself rather than in terms of the typical occupant of the role. According to the occupant functionalist (e.g., Lewis), if neural state N typically plays the role in humans associated with having a desire for coffee, and I am in state N, then I have a desire for coffee even if in my particular case N exhibits none of the causal roles associated with such a desire, e.g., it does not make me more likely to accept or drink a cup of coffee if offered one or to express my desire for coffee if asked if I'd like some. Many functionalists find such claims counterintuitive; Lewis concedes that such cases would indeed be odd but nonetheless argues they would be still be cases of my having coffee desires despite their nonstandard causal profiles. We need not decide the issue here; it is enough to note that differing views can be found in the literature.

The type–type identity theory has enjoyed a recent though modest rebirth of interest (Hill and McLaughlin, 1999). Some philosophers have looked to it as a means of solving or dissolving the supposed explanatory gap that confronts and baffles those who try to explain how and why any given conscious state correlates with or might be realized by a given neural state. These neo-identity theorists argue that because it is an identity that is involved there is no explanatory gap to bridge. There are not two things whose linkage needs to be explained; there is just one thing, and it like everything else is necessarily identical with itself and not with anything else. If Brian's pain just *is* a certain pattern of brain activity in the *identity* sense of 'is', then there is no gap to be closed any more than there is any case of identity. If there is just one thing, then as a matter of simple logic it's the same as itself. Some have complained that the explanatory lacuna merely reappears as the unsatisfied demand for some account of how the mental and physical pathways might converge on a common referent, and that complaint seems justified. Nonetheless, it's important to acknowledge that identity versions of reductive physicalism, though far less popular than a few decades back, are still alive and being actively defended.

Composition. One seemingly plausible alternative to identity is composition. If mental things (e.g., minds) are *composed* entirely of physical parts, might that not suffice to justify the reductionist claim that they reduce to the physical or are all 'just physical'. If all of a thing's parts are physical, can the reductionist not say

that it 'contains nothing over and above the physical'? Moreover, composition is distinct from identity and has a different logical status that easily accommodates the sort of multiple realizability objections raised against identity theory. Higher-level objects can outlive the components of which they are at a given time composed. A marching band which gradually changes its membership over time can continue to exist even after one hundred per cent of its original members have retired; identity of the band itself is not dependent on the sameness of its underlying composition. So, too, it might seem a given mental state might persist through changes in its underlying neural composition. There may be good mentalistic reasons for regarding the memory that I have on Tuesday of the film I saw on Sunday as the same mental state as the memory that I had of it on Monday, even if the two differ nontrivially in their neural components. Brains and the patterned information they encode seem to be quite neurally dynamic; retaining or recalling a given memory need not rely upon the very same ensemble of neuron activations.

Despite its obvious attractions, composition will not suffice to ground reductive physicalism. Indeed it is compatible with various forms of property dualism, including both fundamental and emergent property dualism. To say that a thing is composed *entirely of physical parts* is not the same as saying that *all its parts are entirely physical*; to assert the former is to say only that all its parts have physical properties, but the second asserts as well that those parts have *only* physical properties. Without the second stronger assertion, nothing excludes the possibility of nonphysical mental properties either at the level of the parts or at the level of the whole (emergent property dualism), and the appeal to composition *per se* will not suffice to support that stronger assertion.

Supervenience. Some philosophers in recent years have appealed to the notion of supervenience as a way of getting beyond mere composition and reductively linking mental and physical properties themselves while still stopping short of strict identity (Davidson, 1970; Kim, 1982). The issue is complicated by the lack of any philosophical consensus about the identity and individuation conditions for properties; there is no dominant view in the field about the metaphysical status of properties and the conditions under which one property is the same as another. Nonetheless, various attempts have been made to analyse the relation between mental and physical properties in ways that might legitimate physicalism without invoking a strict type–type identity of properties.

Supervenience, which involves the dependence of one set of properties on another, was first proposed early in the twentieth century as a way of explaining the relation between normative properties such as moral and aesthetic properties and their non-normative bases (Moore, 1902). The basic idea is that one set of properties (X-properties) supervenes on another (Y-properties) such that there can be no X-differences without Y-differences, or to put the point the other way round, any two things sharing all their Y-properties must also be alike in all X-properties. For example, the beauty of a painting may not be identical with any of its strictly physical properties such as its distribution of pigments on the surface of the canvas, but any other painting that shared all the physical properties of the first would also have to share its aesthetic properties. If the first were

sublimely beautiful, so too would be the second. In the mind/body domain, the basic view can be conveyed by the slogan, 'No mental difference without a physical difference' (Davidson, 1970; Kim, 1982). Although supervenience theorists have not generally labelled themselves as reductionists, the dependence relation provides one way to cash out the basic reductionist idea that 'Xs are just Ys'.

Although supervenience enjoyed a brief period of intense interest as a possible way of making sense of ontological physicalism, it has now generally fallen out of favour. Even Jaegwon Kim, who played the largest role in bringing the notion to the centre of discussion (1982; 1993a) has acknowledged more recently (1999) that supervenience is too weak a relation to validate physicalism and is *a fortiori* an inadequate way of analysing the concept of reduction. Kim, for example, has conceded that supervenience is compatible with both property dualism and dual aspect theory. Even if the mental and physical realms were distinct and separate, supervenience could still hold as long as there were invariant correlations between the two — whether underwritten by natural law (*nomic supervenience*) or some stronger metaphysical link (*metaphysical supervenience*). Indeed, the first sort of correlation would hold in the world of a classical interactive dualist like Descartes and the latter in the world of a panpsychic monist like Spinoza. If supervenience is compatible with such explicitly nonphysicalist views, it seems unlikely to provide an adequate account of physicalism and even less of how the mental might reduce to the physical. Some (e.g., Chalmers, 1996) have replied that what is needed is a relation of *logical supervenience*, according to which as a matter of logic it is impossible that mental and physical properties might independently vary. But then the question quickly arises of what sort of link short of identity might underwrite such a logically necessary link. Thus, if an adequate answer is to be found, it looks like the real explanatory work will have to be done by the story of that underlying relation (whatever it might be) rather than by the notion of supervenience *per se*.

Realization. One way of spelling out that underlying story might be in terms of realization. The multiple realizability of mental states was used as an objection against type–type identity theory (Putnam, 1972), but it has also provided the basis for a positive view of the psycho–physical link that might suffice as a form of ontological reduction. Realization is especially attractive to functionalists. Two systems can manifest the same functional property even if it is realized by different structures in the two cases. This holds both for natural biological functions and artefactual ones. Nature can typically build a membrane with a given permeability profile in more than one way, and engineers can design a signal amplifier with the same input/output function using many different hardware setups. However, as an ontological matter, the given functional property is fully realized in each case by its underlying physical components and their mode of composition. Once you've fixed all the facts about the structures and processes at the physical level, the facts about the functional properties follow automatically.

Realization appeals to those who favour a mind/computer analogy (Putnam, 1972; Fodor, 1981) since one and the same software or computational processes can be realized by many different types of hardware. However, realization is also

invoked by some philosophers like John Searle (1992), who are anti-computationalist and explicitly nonfunctionalist about the mental. Searle denies that mental properties can be functionally analysed because of what he regards as their irreducible first-person intrinsic features, but he nonetheless classifies his view as physicalist by appeal to his slogan that mental properties are 'caused by and realized in' the physical processes of the brain. For Searle, the realization claim is essential to avoid property dualism, which would remain an option if he claimed merely that mental properties were caused by physical brain processes. But with the addition of the realization requirement, he regards himself as having shown mental states as metaphysically no more problematic than the liquidity of room-temperature water, which is analogously *caused by and realized in* interactive collections of H_2O molecules.

Realization plays a role as well in many versions of nonreductive physicalism. In her attempt to combine ontological physicalism with a denial of representational reduction (ONT-Reduction & Not REP-Reduction), the nonreductivist most often appeals to realization to secure her credentials as an ontological physicalist (Van Gulick, 1992). Can she succeed? Can she give an account of a realization that is strong enough to vindicate physicalism, yet consistent with a robust denial of REP-reduction? We will get to that question below, but first we need to shift our attention from the ontological side (ONT-Reduction) and consider the equally diverse family of relations that qualify on the representational side as kinds of REP-reduction.

In a case of REP-reduction, one set of representational items is reduced to another. In answer to Question 1, we noted above that REP-reductive relations might hold among at least four different kinds of representational items: concepts, theories, models and frameworks. Those four do not exhaust the relevant options, but for present purposes we can restrict ourselves to them. What then of the representational version of Question 2:

> Question of the *representational link:* How must things be related for one to representationally reduce to the other?

There are a diversity of answers given in the literature, some of which make sense only with respect to certain kinds of items in the representational domain; some relations, such as derivability, might make sense as relations between theories but not between models or among concepts. However, certain commonalities run through the family of REP-reductive relations. They all involve some sort of *intentional equivalence*, i.e., some correspondence in terms of what they can or do represent as opposed to a mere correspondence in their form or intrinsic properties. The basic idea is that one representational item (or set of items) reduces to another just if the first is linked to the second in terms of what it can or does say about the world and its features. Thus, REP-Reductive relations generally concern either the comparative expressive powers of representational items (what they *can say*) or correlations in their assertoric content (what they *do say).* To invoke a hoary and controversial traditional division, representational reductions

turn on either relations of meaning or relations of truth. Most of the specific variants of REP-Reduction fall into one of five general categories:

- Replacement ,
- Theoretical–Derivational (Logical Empiricist),
- *A priori* Conceptual Necessitation,
- Expressive Equivalence (two-term semantic relation),
- Teleo–Pragmatic Equivalence (n-term pragmatic relation).

These alternatives have been extensively debated in the literature, but I will say only enough about each to indicate the range of options that are active in the field.

Replacement. The analogue on the representational side of elimination on the ontological side would be replacement. Our prior ways of describing and conceptualizing the world might drop out of use and be superseded by newer, more adequate ways of representing reality. For example, many of our mentalistic concepts might turn out not to do a good job of characterizing the aspects of the world at which they were directed, as has happened with demon concepts. If so, future science might develop alternative concepts and theoretical resources that would more accurately and effectively represent reality. If adopting those newer systems of representing should lead us to drop our former mentalistic outlook, then in an extreme sense our mentalistic way of speaking and thinking might be regarded as having been reduced to its representationally superior replacement. However, most notions of REP-Reduction to be found in the literature are more conservative and involve preserving more of the truth or expressive content of the reduced theory.

Theoretical–Derivational. The classic notion of reduction in terms of theoretical derivation, as found in Kemeny and Oppenheim (1956) or in Ernest Nagel's classic treatment (1961), descends from the logical empiricist view of theories as interpreted formal calculi statable within the resources of formal symbolic logic. Given the axioms or laws of such a theory together with a formally statable set of actual conditions, one can derive all its consequences, observational and otherwise, by working out its formal implications. Thus, if one such theory T1 could be logically derived from another, T2, then everything T1 says about the world would be captured by T2, and T1 could be said to reduce (REP-reduce) to T2. Because the theory to be reduced, T1, normally contains terms and predicates that do not occur in the reducing theory T2, the derivation also requires some bridge laws or bridge principles to connect the vocabularies of the two theories. These may take the form of strict biconditionals linking terms in the two theories, and when they do, such biconditionals may underwrite an ontological identity claim. If a gas has a given temperature when and only when its molecules have a given average kinetic energy, then we may go on to infer that temperature just *is* average kinetic energy, and that heat is *identical with* molecular motion. However, the relevant bridge principles can also take other forms; they need not be strict biconditionals. All that is required is enough of a link between the vocabularies of the two theories to support the necessary derivation.

One other caveat is in order. In most cases what is derived is not strictly speaking the original reduced theory but an image of that theory within the reducing theory, and that image is typically only a close approximation of the original rather than a precise analogue (Churchland, 1985). For example, in the paradigm case of the reduction of classical thermodynamics to statistical mechanics, the image of the classical laws within the statistical domain allows for possible though extremely improbable deviations from the classical laws. Thus, as a strict matter of logic, those prior laws are not derived. However, we accept that the classical laws were not strictly speaking true, and the match between the original laws and their statistical analogues is so close that we accept the former as having been reduced to the latter.

Were we to try to apply the theoretical–derivational model of REP-reduction to the mind/body case, we would need to find a set of bridge principles that allowed us to derive all the truths of our mentalistic theories of consciousness from the laws and statements of the relevant reducing theory (theories) whatever they might be: neurophysiological, computational, quantum mechanical or otherwise. As in the thermodynamics case, the derived result might not be an exact analogue, but it would have to be a close enough approximation to our original pre-reductive mentalistic theory to justify the claim of REP-reduction by theoretical derivation.

Such a prospect might have once been viewed as the likely result of eventual scientific progress; logical empiricist views of the unity of science envisioned such an eventual formal integration of our scientific representation of reality (Oppenheim and Putnam, 1958). At present, far fewer philosophers expect such derivations to be produced even in the long term. Ontological dualists believe that our mental and physical theories describe different domains, and thus they expect no reduction of the theories that describe them. But REP-reduction is deemed unlikely also by many physicalists, especially those who accept some form of nonreductive physicalism. Unlike the dualists, they believe there is just one domain of reality which at some level is correctly represented by physics, but they also believe that adequately representing all the complex features that that reality exhibits at its many different levels requires the use of a diversity of theoretical and representational resources beyond those provided by the formal structures of physical science *per se* (Fodor, 1974; Boyd, 1980; Garfinkel, 1981; Van Gulick, 1992). We will inquire further below about why nonreductive physicalists believe that, and about whether such a view can be consistently combined with ontological physicalism. For the present we need only note that for a diversity of reasons, theoretical REP-reduction by derivation is rejected in the mind/body case by many but not all current philosophers.

A Priori Conceptual Necessitation. Though many reject theoretical reduction as too strong a link, other philosophers regard it as too weak for adequately REP-reducing mind to body. In particular, they demand that the bridge principles not merely support the derivation of the reduced theory but do so by linking the two theories via necessary and *a priori* conceptual links or inter-theoretical definitions (Levine, 1983; 1993; Chalmers, 1996). They would not count a derivation

based upon mere empirically established links or biconditionals as a successful reduction, since such empirical bridge principles might only describe the correlations between properties in distinct and separate mental and physical domains. This alleged weakness of the derivational account of reduction is the representational analogue of the ontological faults charged against supervenience. Lawlike covariance, though sufficient to support nomic supervenience on the ontological side and inter-theoretical derivation on the representational, nonetheless seems inadequate to vindicate physicalism or reductionism because it is compatible with both property dualism and dual aspect theory as long as the mental and physical domains are lawfully linked; something which even a Cartesian dualist accepts. Proponents of the *a priori* view argue that nothing short of logically sufficient conceptual connections will suffice for one theory to reduce another (Chalmers, 1996). They often couple that claim with a demand for reductive explanation, i.e., for an explanatory account that lets us see in a conceptually necessary way how the conditions described by the reducing theory must as a matter of logic alone guarantee the satisfaction of those described by the reduced theory. They appeal to models such as that of the liquidity of water. The explanation in such cases supposedly includes two components:

- first, an analysis of the concept to be reduced (e.g., liquidity) in terms of some set of typically functional conditions, and
- second, an account of how those conditions would as a matter of mere logic be satisfied by any underlying system meeting the conditions described by the reducing theory (the micro-interactions of the collection of room-temperature H_2O molecules).

The 'apriorists' contend that nothing less would suffice as a theoretical reduction of the mental to the physical.

Once again, the issues are numerous and debates in the literature extensive, but we need only note the controversial nature of the apriorists' claim. Though the *a priori* view has intuitive appeal and its share of supporters, it also has a host of critics (Block and Stalnaker, 1999; Van Gulick, 1999; Yablo, 1999), some of whom plausibly charge it with setting up — or at least implying — a false dilemma. Most physicalists would agree that mere brute fact mental–physical correlations unsupported by appeal to any underlying explanation of why they were so linked would fall short of providing a reduction even if they sufficed as a bridge to derive our mental theories from our physical ones. Indeed, right at the start of contemporary physicalism, the philosopher Herbert Fiegl (1958) disparaged such unsupported links as 'nomological danglers'. However, demanding that the bridge principles provide logically necessary and *a priori* conceptual links between the two domains seems to swing to the opposite extreme. Surely there must be intermediate cases that involve explanatory rather than merely brute links, but that nonetheless fall short of the apriorists' radical requirement for strict logical entailments. There is no consensus about what would count as an adequate explanation of the psycho–physical link and about what sorts of bridge principles would suffice for a satisfactory theoretical derivation, but many

physicalist believe that the answer lies somewhere between the two extremes (Van Gulick, 1992; Kim, 1999).

Expressive Equivalence. The two versions of REP-Reduction considered thus far both concern what we might call truth preservation, i.e., in both cases everything the reduced (mental) theory says about the world is also asserted by the reducing (physical) theory combined with the requisite sorts of bridge principles. However, REP-Reduction might be viewed not as a matter of equivalence in what *is said* about the world, but as merely a matter of preserving the expressive range of what one *can say.* One representational system, R1, might be regarded as reducible to another, R2, as long as every state of affairs representable by R1 could also be represented by R2. The expressive range of R1 would be entirely contained within that of the reducing system R2.

Though expressive equivalence seems to set a weaker requirement than the derivational or apriorist versions of REP-Reduction, it is still far from obvious that such a criterion can be met. Many philosophers, dualist and otherwise, have appealed in particular to the alleged special nature of first-person phenomenal concepts and the sorts of experiential facts that we can supposedly know or understand only through their use. Drawing on an old empiricist intuition that goes back at least to John Locke (1688) in the seventeenth century, contemporary philosophers such as Thomas Nagel (1974) and Frank Jackson (1982; 1986) have argued that there are facts about consciousness that can be adequately known or understood only from the first-person experiential perspective; for example, one can fully understand what it's like to taste a pineapple only if one has oneself had such an experience. (Although Jackson himself has recently [1998] changed his mind on this issue, many others continue to invoke his original position and have not followed him in his reversal of opinion.) Thus, some philosophers claim that physical theory with its reliance on third-person concepts can never fully achieve the expressive range of our mental modes of representing, especially those that involve experiential concepts. If there are indeed subjective facts that lie beyond the representational power of physical theory, then it may be impossible to REP-Reduce the mental to the physical even in the weaker sense of expressive equivalence. There may be things that we can say mentalistically about the world that fall outside the range of what can be said using the resources of physical theory. Unsurprisingly, philosophers disagree about whether or not there are such facts (Lewis, 1982; Churchland, 1985; Van Gulick, 1985; 1993a; Levin, 1986; Loar, 1990; Lycan, 1990), but for present purposes we need only note that if there are, then that would seem to preclude REP-Reducing the mental to the physical in the expressive equivalence sense.

Teleo-Pragmatic Equivalence. The expressive equivalence account of REP-Reduction treats representation as primarily a two-term relation between a representation (or set of representations), R, such as words or sentences in a language or theory, and the item (or items), X, represented by R. Moreover, the relation is thought of in terms of familiar semantic notions such as reference and meaning. Thus, if we interpret REP-Reduction as a matter of expressive equivalence, the question becomes: For every representing element of T1 that means M or refers to

an item X, is there a corresponding element or combination of representing elements in T2, call it R*, that has that same meaning or referent? This is certainly one legitimate way to define REP-Reduction, but it may not be the most helpful or revealing. Though representation can be viewed as a two-term relation, doing so involves abstracting away from other significant parameters of representation that are likely to be of importance to understanding the REP-reductive or non-REP-reductive nature of the mind/body link. In particular, what a given representation, R, succeeds in representing is crucially dependent upon the causal structure of the representation-user, U, the social and physical context, C, in which R is applied, and the modes of causal and epistemic access that R affords to U when used in C. Thus, rather than being just a two-term relation between a representation, R, and represented item, X, representation is at least a four-term relation that includes as well a representation user, U, and a context of use, C; nor need we stop at four, additional parameters might easily be added such as the goals or ends toward which the representation is to be applied. Reframing the question of REP-Reduction in terms of this more complex relation, the question becomes as follows: If U can use R from theory T1 in context C to represent X, is there an R* from T2 that U can similarly use in C to represent X, or at least some R† from T2 that U can use in some context C† to represent X? To put the matter less abstractly, the problem is to find a way, if possible, for us to use the contextually embedded resources of the reducing theory to do the equally contextual representational work done by the items in the theory we are trying to reduce. Nor should we ignore the third or pragmatic parameter mentioned above. Success in real-world representation is in large part a practical matter of whether and how fully our attempted representation provides us with practical causal and epistemic access to our intended representational target. A good theory or model succeeds as a representation if it affords us reliable avenues for predicting, manipulating, and causally interacting with the items it aims to represent. This is no less true in the natural realm when one is judging that a given structure in the rat hippocampus serves as allocentric representation of its spatial environment than it is when we judge that an economist's model has succeeded in representing the effect of interest rate changes on housing markets. In both cases, it is the practical access that the model affords to its user in its context of application that justifies us in viewing it as having the representational content that it does (Van Gulick, 1992).

Viewed from this more inclusive contextual and pragmatic perspective, the question of REP-Reduction becomes one about the ability of representation-users to gain the same modes of access with the alleged reducing representations that they do with the representations to be reduced. In the particular mind/body case, this turns in part on our ability (or inability) to use the representational resources of physical theory to replicate the functional interactive profile of the access afforded us by our mentalistic resources including our first-person concepts and theories. Put in this way, REP-Reduction may seem far less plausible; it seems unlikely that we could use physical theory to gain access to our mental states and processes in ways that afford us the same understanding that we achieve through

our first-person and introspective modes of representation. The differences in the contexts of use are so great that they alone seem sufficient to make such an equivalence for all practical purposes impossible. Many of our first-person concepts are so directly embedded within our intra-mental processes of self-regulation, self-monitoring and self-modulation that it is difficult to see how any third-person system of concepts provided by physical theory could achieve a pragmatic profile sufficiently similar to support a claim of REP-Reduction. Indeed, it is for just that reason that philosophers who adopt a pragmatic view of representation typically also deny the possibility of REP-Reduction and are in that sense nonreductionists (Putnam, 1978; Garfinkel, 1982; Van Gulick, 1992).

Though much more could be said about the many varieties of ontological and representation reduction and their respective faults and merits, I hope at least to have surveyed the main versions of each as graphically summarized in figure 3.

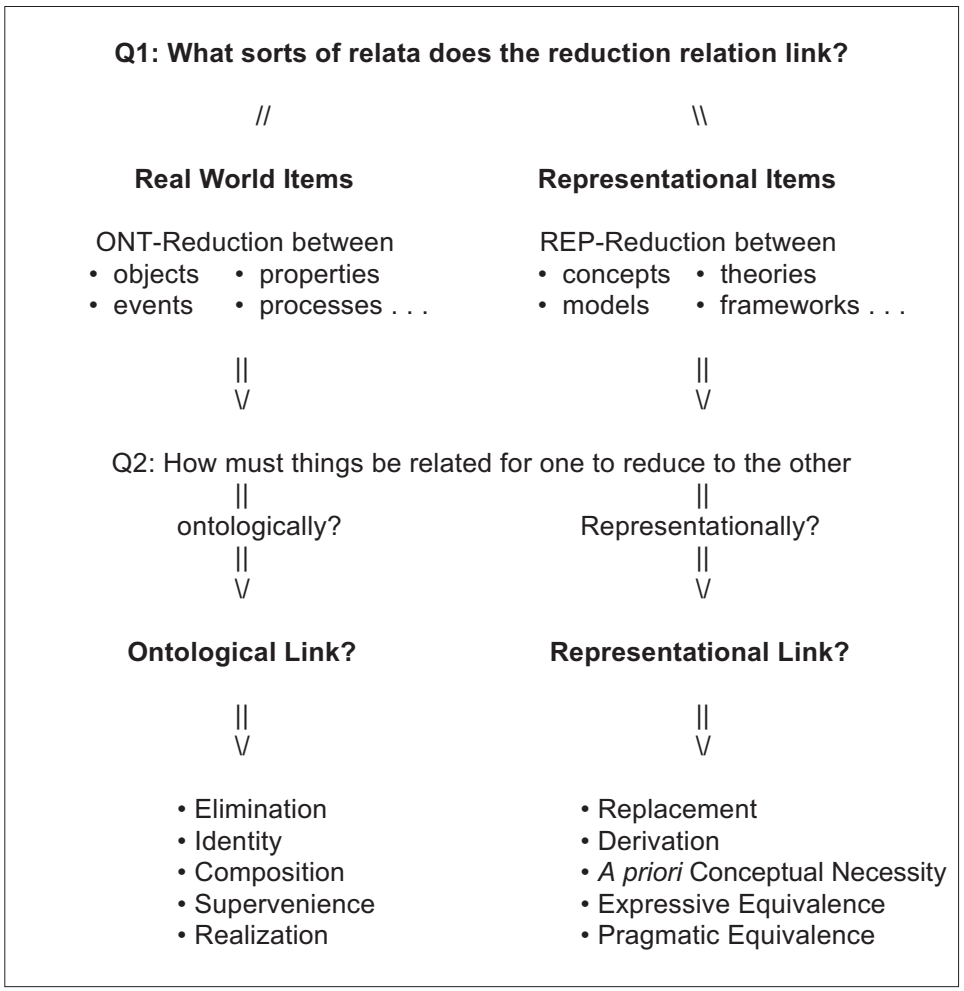

Figure 3

III: The Varieties of Emergence

We must thus turn our attention to the notion of emergence, which like reduction gets interpreted in diverse ways in the mind/body literature (Searle, 1992; Hasker, 1999; Silberstein and McGeever, 1999). Again, my aim will be to survey the main variants. Only when that has been done can we investigate the relations among the many members of these two diverse families of concepts.

The basic idea of emergence is more or less the converse of that associated with reduction. If the core idea of reduction is that Xs are 'nothing more than Ys' or 'just special sorts of Ys', then the core idea of emergence is that 'Xs are more than just Ys' and that 'Xs are something over and above Ys'. Though the emergent features of a whole or complex are not completely independent of those of its parts since they 'emerge from' those parts, the notion of emergence nonetheless implies that in some significant and novel way they *go beyond* the features of those parts. As we will see, there are many senses in which system's features might be said to emerge, some of which are quite modest and unproblematic (Searle, 1992) and others which are more radical and controversial (Hasker, 1999).

The varieties of emergence can be divided into several groups along lines that are similar in at least some respects to the divisions among the types of reduction. For example, emergence relations might be viewed either as objective metaphysical relations holding among real-world items such as properties, or they might be construed as partly epistemic relations that appeal in part to what we as cognitive agents can explain or understand about such links.

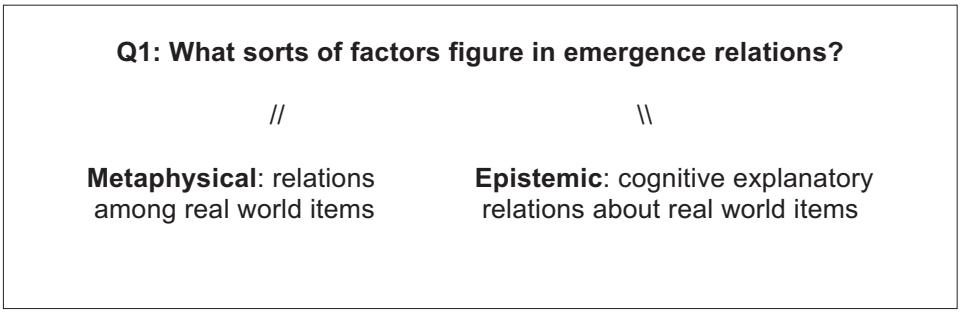

Figure 4

Relations of the first sort are objective in the sense that they concern the links or lack of links between real items such as properties, independent of any considerations about what we humans, as epistemic subjects, can or cannot explain or understand about them. Relations of the second sort are epistemic, and in a sense subjective, because they turn crucially on our abilities or inabilities to comprehend or explicate the nature of the links or dependencies among real-world items rather than just on those links alone. The two sorts of notions often get run together in the discussion of emergence, but it is important to keep them distinct just as with the ontological and representational notions of reduction discussed above (Silberstein and McGeever, 1999).

On the objective side, two main classes of emergents can be distinguished: properties and causal powers or forces. The distinction between the two is not sharp and involves possible overlaps, especially if one individuates properties in terms of their causal profiles. Nonetheless, the issue of emergent causation is critical, so it's worth distinguishing at least initially between emergent properties and emergent powers, even if the line subsequently blurs a bit. Within each of the two classes, there is a continuum of cases running from the extremely modest to the extremely radical. The former involve emergent features, whether properties or powers, that are very similar in nature to the features from which they emerge, whereas the emergents in the latter sorts of cases are most unlike their nonemergent bases. Although the cases differ by many variations of degree, we can give some sense of their range by dividing the cases into three rough categories of increasing radicality which we can label: specific-value emergence, modest-kind emergence, and radical-kind emergence.

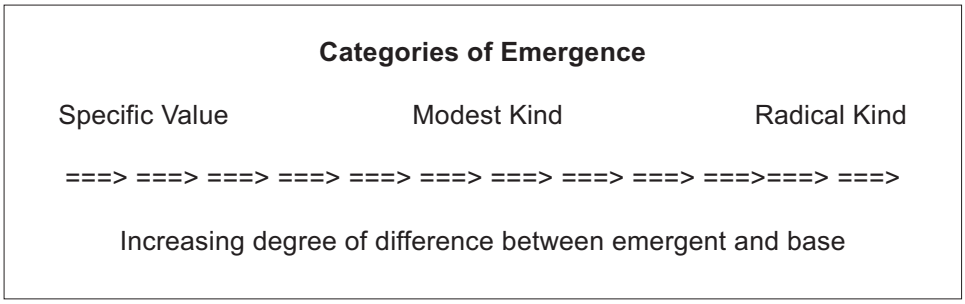

Figure 5

We can define the three roughly as follows.

- Specific Value Emergence. The whole and its parts have features of the *same kind*, but have different *specific subtypes or values* of that kind. For example, a bronze statue has a given mass as does each of the molecular parts of which it is composed, but the mass of the whole is different in value from that of any of its proper material parts.
- Modest Kind Emergence. The whole has features that are *different in kind* from those of its parts (or alternatively that *could* be had by its parts). For example, a piece of cloth might be purple in hue even though none of the molecules that make up its surface could be said to be purple. Or a mouse might be alive even if none of its parts (or at least none of its subcellular parts) were alive.
- Radical Kind Emergence. The whole has features that are both
  1. different in kind from those had by its parts, and
  2. of a kind whose nature and existence is not necessitated by the features of its parts, their mode of combination and the law-like regularities governing the features of its parts.

Whether or not there are any cases of radical-kind emergence is controversial. Physicalists who would readily concede the reality of specific-value emergence and modest-kind emergence would most likely deny there are cases of the more extreme sort. Accepting radical-kind emergence would be conceding that there are real features of the world that exist at the system or composite level that are not determined by the law-like regularities that govern the interactions of the parts of such systems and their features. Doing so would require abandoning the atomistic conception, which is typically embraced by mainstream physicalism. Radical-kind emergence would require giving up at least one of two core principles of atomistic physicalism (AP):

- AP1. The features of macro items are determined by the features of their micro parts plus their mode of combination. (In a slogan: Micro features determine macro features.)
- AP2. The only law-like regularities needed for the determination of macro features by micro features are those that govern the interactions of those micro features in all contexts, systemic or otherwise.

The idea of AP2 is that there are no laws governing micro features in systemic contexts other than those that govern them outside such contexts. The intent is to exclude special laws that come into play only in restricted systemic contexts. Once the micro features and their distribution have been fixed, the micro necessities by themselves suffice to determine the macro properties of the system. No further laws or law-like regularities are needed to necessitate the macro outcomes. Mainstream atomistic physicalism thus includes a commitment to what might be called 'bottom-up determination'. To borrow an example from the philosopher Saul Kripke (1972), once God had done the work of fixing the micro features and laws of the universe, there was no work left to do; in fixing the world's micro nature, He had already determined all its macro properties as well. Or so, at least, atomistic physicalists would claim. As we will see below in discussing the dual revolutions position, some physicalists believe we should give up our commitment to atomism and micro-physical determination in favour of a more holistic view of physical reality; such an alternative though not inconsistent with physicalism *per se* is certainly far from the mainstream view, and it will best to consider it below under the category of other options.

If we cross-pair our three rough divisions along the continuum of unlikeness between emergents and bases with our two classes of objective emergents (properties and powers), we then get six versions of metaphysical emergence as shown in figure 6.

The notion that causal powers might exhibit radical-kind emergence merits special attention since it poses perhaps the greatest threat to physicalism. If wholes or systems could have causal powers that were radically emergent from the powers of their parts in the sense that those system-level powers were not determined by the laws governing the powers of their parts, then that would seem to imply the existence of powers that could override or violate the laws governing the powers of the parts; i.e., genuine cases of what is called 'downward causation'

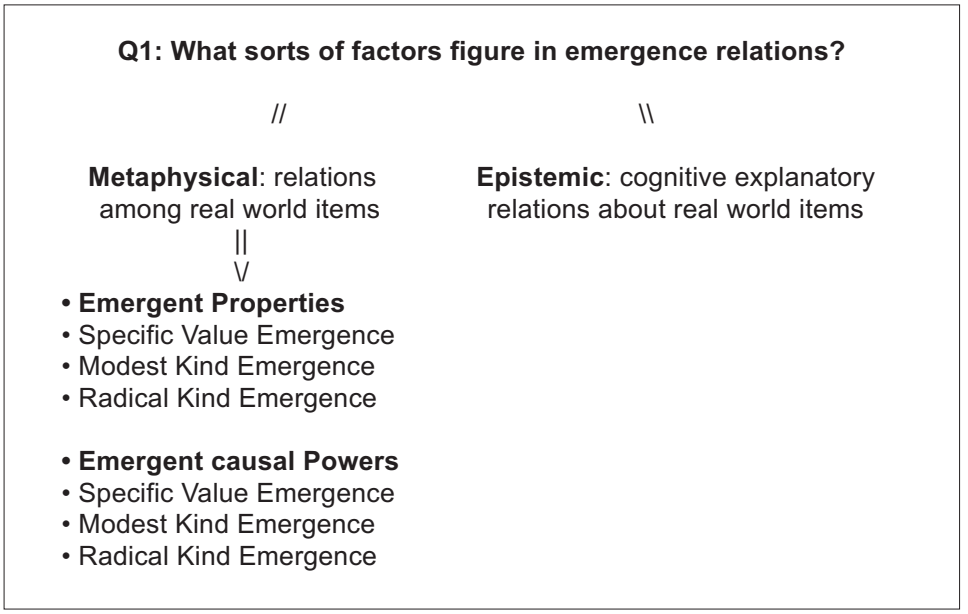

Figure 6

(Sperry, 1983; 1991; Kim, 1992; 1999; Hasker, 1999) in which the macro powers of the whole 'reach down' and alter the course of events at the micro level from what they would be if determined entirely by the properties and laws at that lower level. All that is needed to get such a result is the relatively uncontroversial claim that macro-level causal powers have micro effects. That in itself need cause no trouble for the physicalist as long as the macro powers are themselves determined by the micro powers (Kim, 1999). But if some macro causal powers were radically emergent, that would free them of determination by the underlying micro powers thus allowing them to alter the course of micro events in ways independent of the micro-level laws. If the physicalist wants to avoid violation of underlying physical laws, she can allow macro-level properties to have micro effects only if those macro powers are themselves constrained by the micro-level laws, which is of course just what the radical emergence of causal powers denies (Hasker, 1999). It is in this respect that radically emergent causal powers would pose such a direct challenge to physicalism, since they would threaten the view of the physical world as a closed causal system; i.e., the idea that nothing outside the physical causally affects the course of physical events (Kim, 1990; 1999). Unsurprisingly, this feature that makes radical causal emergence so threatening to physicalists (e.g., Kim, 1992; 1999) is the very one that make it so attractive to those emergentists like William Hasker (1999) who invoke emergence in support of ontological dualism.

The challenge to those who wish to combine physicalism with a robustly causal version of emergence (Sperry, 1991; Van Gulick, 1992) is to find a way in which higher-order properties can be causally significant without violating the basic causal laws that operate at lower physical levels. On one hand, if they override the micro-physical laws, they threaten physicalism. On the other hand, if the higher-

level laws are merely convenient ways of summarizing complex micro-patterns that arise in special contexts, then whatever practical cognitive value such laws may have, they seem to leave the higher-order properties without any real causal work to do (Kim, 1992). One possible solution would focus on the respect in which higher-order patterns might involve the selective activation of lower-order causal powers. Micro-properties that were causally irrelevant in most configurations, e.g., because their random actions cancelled out and had no significant overall effect, might exert a powerful causal influence in a small range of cases involving higher-level patterns that brought those micro-powers into a coherent mode of action making a major difference to the overall operation of the system (Van Gulick, 1993b). For example, the magnetic fields of molecules can be ignored in explaining most materials since the are randomly aligned. However, when they are coherently oriented in a magnet, those micro-properties are crucial to understanding the causal powers and activity of the system that contains them. Such selectional models need pose no problem for atomistic physicalism since they involve no violation of underlying micro-physical causal laws. However, whether they accord a sufficiently potent form of causal efficacy to macro or system-level properties to justify a claim of emergent causality is open to debate. The selective activation model nonetheless provides at least an example of how one might try to find a way of reconciling physicalism and causal emergence.

Having given at least six rough options on the metaphysical side, let us turn our attention to epistemic notions of emergence. What makes all such notions epistemic is that they involve some respect in which we are unable to predict, explain, or understand the features of wholes or systems by appeal to the features of their parts. Emergence in this sense is thus at least in part subjective, i.e., a matter of our cognitive and explanatory capacities and limits rather than just a matter of relations between objective items as in the metaphysical cases. For present purposes, it will probably suffice to distinguish two versions of epistemic emergence, which focus respectively on different cognitive abilities — the first on prediction and explanation and the second more on representation and understanding. Once again, the lines between the two are not sharp, but worth distinguishing.

Predictive/Explanatory Emergence: Wholes (systems) have features that cannot be explained or predicted from the features of their parts, their mode of combination, and the laws governing their behaviour.

Representational/Cognitive Emergence: Wholes (systems) exhibit features, patterns or regularities that cannot be represented (understood) using the theoretical and representational resources adequate for describing and understanding the features and regularities of their parts.

Both versions of epistemic emergence come in weak or restricted forms and in strong or unrestricted forms. On one hand, the relevant cognitive inability to explain or represent might be a restricted fact about our specifically human limitations, or even more restrictively about our present state of theorizing and scientific progress. Alternatively, the cognitive inability might concern a more general limit that applies universally to all cognitive agents or to all those in some broad

category of which we humans are just one example among many. The philosopher Colin McGinn, for example, has argued that humans lack the ability to form the sorts of concepts needed to make the psycho–physical link intelligible (McGinn, 1989; 1991). Thus, he believes it will ever remain a mystery to us, even though he accepts consciousness as an aspect of physical reality and allows that cognizers with concept-forming abilities quite different from our own may be able to comprehend the link in intuitively satisfying ways. It's not that we humans are not smart enough; it is that we have the wrong sorts of minds for solving the psycho–physical puzzle. From McGinn's perspective, consciousness is epistemically emergent in the explanatory sense, at least relative to humans and other cognitive agents with our sorts of conceptual capacities.

We can summarize our quick survey of the varieties of emergence in figure 7.

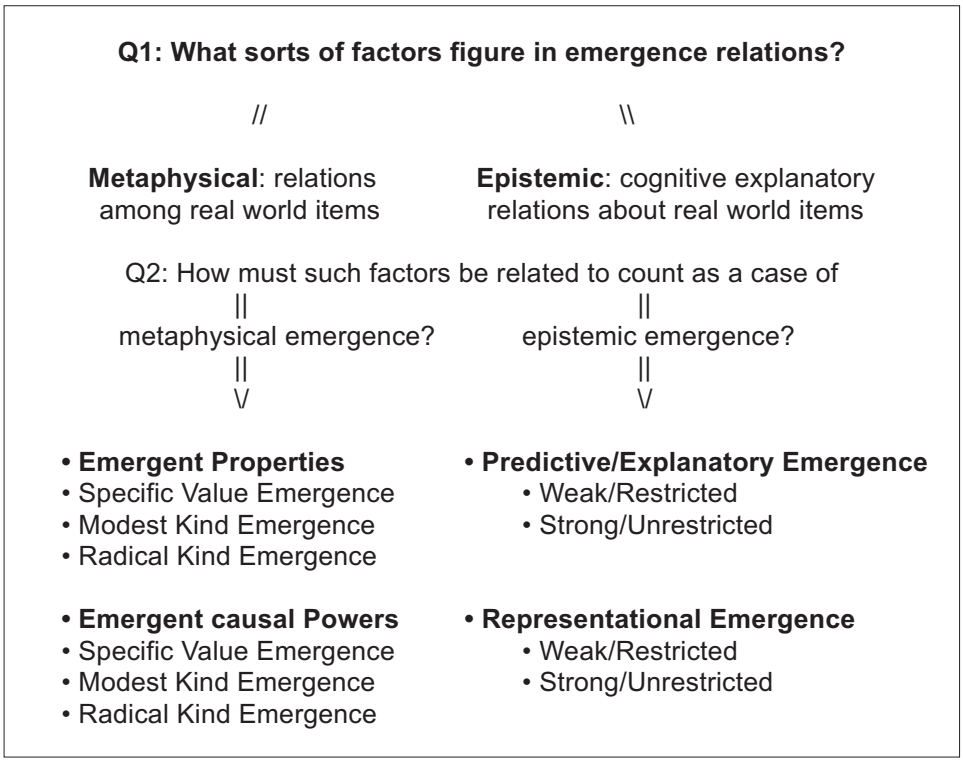

Figure 7

IV: Other Recent Mind–Body Options

Our taxonomizing survey has thus distinguished among ten varieties of emergence and at least ten versions of reduction. Even so, it probably does not capture every interesting variant to be found in the literature, but I hope it includes all the major ones and gives a fair sense of the range of options being discussed in both families. There are, of course, current mind/body views that seem to fall into neither the reductive nor the emergent category. Though appearances may mislead

us in some cases, as we will see below in section V, there are current options that are generally regarded as in neither family. Most prominently there is the sort of nonreductive physicalism discussed above which is probably the closest to the mainstream view within the philosophic community (Davidson, 1970; Putnam, 1972; 1978; Fodor, 1974; Boyd, 1980; Searle, 1992; Van Gulick, 1992). As noted, it aims to combine some form of ontological physicalism, typically framed in terms of realization, with a rejection of most forms of representational reduction (ONT-Reduction & Not REP-Reduction). To its supporters, this offers the best of both worlds, while its critics (Kim, 1989; Chalmers, 1996) see it as cheating or attempting (unsuccessfully) to pass as a member of the physicalist club without paying one's mandatory dues. Relevant to our current concerns, some of its critics claim it cannot help sliding into an emergent dualism of a particular problematic sort that entails radical causal emergence (Kim, 1992). Were that so, it would indeed be hard-pressed to defend its alleged physicalist status, but it is far from clear that nonreductive physicalism is forced into any such dire commitments.

Current forms of nonreductive physicalism trace back to the mid 1970s and the work of the philosophers Donald Davidson (1970; 1974), Jerry Fodor (1974), Hilary Putnam (1972) and Richard Boyd (1980) among others, which supplanted prior logical empiricist views about the unity of science and the expectation of tight inter-theoretical links and definitions among our various ways of describing and explaining the many levels and aspects of reality. Fodor in particular was concerned to validate what he called the 'autonomy of the special sciences'. Although everything in the world might be ultimately physical, Fodor argued that the nonphysical sciences — be they biology, economics or psychology — provided us with means of describing, explaining, predicting and manipulating the world that were unavailable using the resources of the physical sciences. The old unity of science view (Oppenheim and Putnam, 1958), which held that all true theories must ultimately be translatable into the language of physics, was rejected as a form of conceptual or representational imperialism. Putnam (1978) drew heavily on the pragmatic aspects of representation and explanation and on a practical view of theories as cognitive tools. Whether an all-knowing God might or might not be able to use purely physical descriptions to comprehend all the complexities of the world was of little relevance to our theoretical options as the contextually situated and cognitively limited agents that we are. For all practical purposes, the explanations and understanding available to us through the use of the special sciences are simply not accessible through the use of physical theory alone (Garfinkel, 1981; Van Gulick, 1992). And when one is engaged in the real practice of science, what matters is what is possible in practice by us rather than what an omniscient deity might be capable of doing.

The view obviously has its attraction and fits well with the pluralistic *Zeitgeist* of our time. The challenge has been to show that one can 'eat one's pluralist cake' while still remaining robustly physicalist at the ontological level. Being a nonreductive physicalist myself, I believe the two aspects of the view can be consistently combined, and I have argued for such a view elsewhere (Van Gulick, 1992). However, critics from both left (dualists such as Chalmers, 1996; Hasker,

1999) and right (reductive physicalists such as Kim, 1989) obviously disagree, and debate continues.

Among other recent mind/body options, at least four need to be mentioned: fundamental property dualism, pan (proto-) psychism, dual-aspect monism and what we may call the 'multi-revolutions view'. In the Kuhnian context of 'extraordinary science' these latter views move farther from the normal paradigm and more radically loosen the rules of what might count as a resolution of the mind/body anomaly. Although I list them as four separate categories, the lines between them blur somewhat and just where a specific theory falls is not always clear. With that caveat, let me say just a bit about each.

Fundamental Dualism. All forms of dualism recognize two distinct and separate ontological domains (the mental and the physical). Though contemporary dualism more often appeals to a duality of properties, historically most dualists were committed to a duality of substances and the latter view still has present supporters.

We have already discussed emergent property dualists. Though they treat mental properties as distinct from physical ones, they do not regard mental properties as fundamental. On the emergent property dualist view, mental properties are something over and above their physical bases, but they are not fundamental in so far as they owe their existence to their emergence from their nonmental physical bases. They are *more than* their bases, but they are nonetheless in some way *dependent upon* them. By contrast, fundamental property dualism gives basic mental properties the same bedrock foundational status as our fundamental physical forces (Chalmers, 1996). Just as the nuclear force and the electromagnetic force are equally primitive features of physical reality according to the so-called standard theory, so too are basic mental and physical properties according to the fundamental property dualist. Thus, any lawlike links between them would be rock-bottom as well and not open to explanation in terms of more basic underlying regularities. If so, we would expect mind/body explanation to terminate at a core of nomically primitive psycho–physical laws, which we would just have to accept as the way things work in our world: 'No more questions please!'

It is certainly a consistent option within the space of possible views, and not open to *a priori* refutation. Yet it seems explanatorily unsatisfying and empirically at odds with our evidence from other cases about how properties at different levels of our world relate (McGinn, 1991). If chemical properties, biological properties, and economic properties are not fundamental, does that not give us at least some good reason for believing that mental properties are probably not either? The property dualist nonetheless believes that there are good reasons for regarding mental properties as a special case (Chalmers, 1996) and debate continues.

Fundamental substance dualists regard minds as independently existing substances or things, distinct from any physical systems with which they may interact. Classical Cartesian dualism treated minds and bodies as distinct substances with their own respective essences. Both popular (i.e., everyday) dualism and traditional religious (Christian) dualism regard souls or minds as distinct

non-physical substances. In recent years the philosophers John Foster and Richard Swinburne have both argued for neo-Cartesian forms of dualism. Their arguments proceed from diverse premises. Foster (1991) relies essentially upon the supposed inadequacy of all forms of reductionism to establish the nonphysicality of mental properties. He goes on to argue that nothing other than a wholly non-physical subject could be the bearer of such irreducible intrinsic mental properties. In effect, he argues from property dualism to substance dualism. Swinburne (1986) appeals to the necessity of a soul to provide an adequate basis for personal identity through time and change, especially for what he regards as the coherent possibility of the persistence and continuity of the self through the loss of one's body and the conceivability of disembodied existence.

All forms of fundamental dualism — whether property or substance — must confront the causal interaction problem that has bedevilled dualism at least since the days of Descartes' critics in the seventeenth century. We seem to have abundant observational evidence of two-way causal interactions between minds and bodies. But if mental properties are fundamentally distinct from the physical, then they could not affect the course of physical events without violating the causal closure of the physical (Kim, 1990; 1999). Thus, the fundamental property dualist seems forced to chose between two unsatisfactory alternatives: giving up the causal closure of the physical or regarding mental properties as epiphenomenal (at least qua the physical domain) despite the enormous body of evidence that appears to show otherwise. Physicalists understandably find neither option attractive. However, many dualists are willing to deny the causal closure of the physical. However much the view may conflict with our mainstream views about the causal closure of the physical world, future evidence could empirically support some form of fundamental dualism, and so it stays on the table as a possibility, whether likely or not.

Pan (Proto-) Psychism. If it is impossible, as some claim, to build a mind or consciousness out of purely nonmental parts, then perhaps we must view the parts as themselves in some way having a mental or at least proto-mental aspect (Nagel, 1979; Chalmers, 1996). That is just what pan (proto-) psychics claim; everything real has a mental (or proto-mental) aspect down to the smallest units of physical reality, though it is difficult to understand in what sense an atom or electron could have features that were even in the most remote sense mental. Arguments for the view are typically indirect and rely on the assumption that conscious minds cannot be constructed out of totally non-mental parts. However, panpsychism, proto or otherwise, is generally very short on details about how mentality might be universally present through out physical reality. The idea of proto-psychic features also confronts a dilemma: the more we view them as *like* familiar mental properties, the more implausible it is to suppose that they could be universally present in simple physical components — how could a molecule be aware, or how could an atom experience red or feel pain? But, conversely, the more we view them as *unlike* familiar mental features, i.e., the more we emphasize their 'proto-ness', the more difficult it is to see how they might give rise to consciousness. The basic explanatory gap just reoccurs at the boundary between

the proto-psychic and the conscious. The viability of the view depends upon giving some plausible positive account of how simple physical components could have mental aspects of a sort apt for producing consciousness, and as yet no such account has been given.

Dual-Aspect Monism. Closely related to the pan-psychism but worth distinguishing, is dual-aspect theory or so-called neutral monism, which asserts that reality is ultimately constituted by a single realm of things and properties (hence the monism) which are neither mental nor physical. This ultimate ('ur') reality manifests itself to us in both physical and mental ways, but is itself more basic than either. Such views might be found historically in Spinoza or Leibniz's theory of monads. Early in the twentieth century, it was advanced on by Bertrand Russell (1921) and various logical empiricists (Schlick, Ayer), and the view still has supporters today (see Strawson, 1994, though he explicitly disavows the 'dual aspect' label in favour of 'agnostic materialism').

Multi-Revolutions View. One of the more extreme responses to the supposed psycho–physical anomaly has been to call for more than one revolution. Proponents of this view, including the physicist Roger Penrose (1989; 1994) and the philosophers Michael Lockwood (1989) and Colin McGinn (1995), have argued that the persistently mysterious nature of the psycho–physical gap gives good reason to believe that we need new ways of conceptualizing and understanding both the nature of the mental and the nature of the physical. Our inability to solve the puzzle of their link results, they say, from the inadequacy of both sides of the equation. McGinn, for example, claims that explaining the link would require both a better understanding of what he calls the hidden nature of consciousness (1991) and a radically different conception of physical space (1995). It is because he doubts that we humans are capable of forming the requisite novel concepts that he takes such a pessimistic view of our human prospects of resolving the anomaly. Lockwood (1989) finds the concept of matter itself deeply problematic and argues for what might be regarded as a dual-aspect view in which matter and mind are more closely integrated at the fundamental level. Penrose finds existing attempts to explain consciousness in terms of physical or algorithmic processes doomed to failure for reasons concerned with the mathematical limits of formal systems; he is equally dissatisfied with the present attempts to integrate our physical theories of the very small and the very large at the interface of quantum mechanics and general relativity. He optimistically hopes for a joint revolution that would address and resolve both puzzles. Some critics regard the search for a 'two-for-one' solution as little more than wishful thinking, but it is one of the features of extraordinary science that people begin to entertain more and more radical alternatives, and sometimes those extreme directions provide the solution, though not often.

A variant of the two-revolutions view argues that the second revolution is needed not in physics but merely in what they regard as the outdated conception of physics held by most philosophers. For example, Michael Silberstein (1999) argues that physics, and especially quantum mechanics, has already rejected the sort of atomistic and localist view assumed by most philosophic parties to the

mind/body debate. Both physicalists and their dualist opponents still typically think of physical explanations in a localist way that requires that all system-level properties be explained as consequences of the properties of the system's parts and mode of combination. Silberstein argues to the contrary that in quantum mechanics, as demonstrated by cases such as those involving entangled particles, systems are regarded as having properties over and above those determined by their parts. One might be a physicalist while rejecting standard atomism. If physics has, in fact, already accepted such a view, which might be regarded as a form of radical metaphysical emergence, then the constraints that limit what can count as a physically acceptable solution to the mind/body problem may indeed be very different from what most philosophers typically assume.

Having completed the survey promised in my title, we can pull all the mind/body options we have discussed into a single diagram which hopefully can serve as a useful pocket map for those travelling through the philosophic terrain. Figures 3 and 7 can be thought of as insets that show more local details in the two ten-member families listed first.

CURRENT OPTIONS ON THE MIND/BODY PROBLEM

•**Reduction**: 10 versions (see *Figure 3*)

•**Emergence**: 10 versions (see *Figure 7*)

•**Other Options**:

Mainstream: Nonreductive physicalism

More radical: Fundamental Dualism
- Property
- Substance

Pan (Proto-) Psychism

Dual Aspect Monism

Multi-Revolutions View

Figure 8

V: Selected Conflicts, Agreements and Other Relations

Having distinguished more than two dozen different options, we can in this final section consider only a few of their possible interrelations (there being 26 x 25 possible two-way cross pairings alone, leaving aside combinations involving more than two views). Since my intent has been to act more as a tour guide rather

than an advocate, I will offer just five general observations aimed at further clarifying the lie of the philosophic landscape. It may contain some paths or links between positions other than those that are normally assumed, and being aware of them should help in navigating the terrain.

With that intent, I offer the following five thoughts on the map and the space of views it aims to depict.

1. *Pay attention to the 'key' — the need for clarity and the avoidance of conflation*. Labels and terms are used with such a diversity of meanings in the mind/body literature that it is absolutely essential that one be clear about what meaning is intended on a given use. Because central terms, such as 'reduction' and 'emergence' have no standard interpretation, one should always make one's own use clear; provide a 'key' for reading the verbal map of your view. And take care in reading others to hear their words as they intended them; otherwise it's all too easy to mislocate them and for the discussion to get lost in a fog of misunderstanding. The general need for clarity is platitudinously obvious, but it's so important in this context that it merits restating nonetheless.

2. *Respect the subjective/objective division.* Perhaps nowhere is the need for clarity greater or the threat of conflation more likely than when one is dealing with subjective and objective versions of some notion. As we saw above, the subjective/objective distinction runs through the space of options dividing both the reduction and emergence families into distinct and separate sections, just as a mountain range might separate a chain of plains or valleys through which it runs.

The reduction region divides into objective relations of ontological reduction such as identity, composition, or realization, and subjective regions of representationally reductive relations such as derivability, conceptual necessitation, or pragmatic equivalence. The parallel division within the emergence region is between objective notions of metaphysical emergence such as modest or radical kind emergence, and subjective notions of epistemic emergence concerning the limits on our cognitive abilities to explain or understand the features of wholes in terms of those of their parts. (N.B. Using the labels 'objective' and 'subjective' here is meant only to draw a distinction between relations among things in the world [objective] and relations among our ways of thinking about or representing things in the world [subjective]. Although those same words are used in many ways in the literature, other meanings should not be read into their use here. In particular, there is no intent to use 'subjective' to imply restriction to a first-person experiential point of view, as when Thomas Nagel classifies facts about experience as subjective because he regards them as fully understandable only from the perspective of those able to have similar experiences themselves. See Nagel, 1974; 1986; Van Gulick, 1985; Lycan, 1990.)

The most common and controversial moves with respect to this division concern attempts to reach ontological conclusions from subjective premises. Facts about our human incapacity to reductively explain how consciousness might be realized by underlying physical processes cannot by themselves justify us in concluding that consciousness is not a physically realized process. The explanatory gaps may reflect subjective limits on our (current) human conceptual or

imaginative capacities rather than any objective divisions in the world. Conceivability arguments for dualism — whether offered by Descartes (1642), Saul Kripke (1972) or more recently by David Chalmers (1996) — are regarded by physicalists as tripping on this mistake. The apparent possibility to conceive of worlds that contain molecule-for-molecule physical duplicates of humans lacking any conscious mental life does not entitle the dualist to conclude that conscious properties are not physical in any ontologically robust sense (e.g., identity or realization). According to the physicalist critics, making such an inference would require us to pass invalidly across the subjective/objective divide, moving from facts about the limits of our concepts to an objective claim about the distinctness and independence of the real-world features to which we refer by use of those concepts. The dualist making such an inference would need to show that the concepts he employed on both the mental and physical sides of his thought experiment were adequate to support such a metaphysical conclusion (Van Gulick, 1999); physicalists doubt the dualist can discharge that burden. In a Shakespearean parody, 'the fault may lie not in the world but in our concepts of it'.

Thus, we should not infer that mental properties are ontologically nonphysical just because we cannot representationally reduce our mental concepts or theories to physical ones. Nor should we conclude that mental properties or powers are metaphysically emergent just because they are subjectively emergent relative to our abilities to explain, predict or understand using resources of physical theory. Additional argument is needed to justify the move from subjective premises to objective conclusion. The dualist champions of conceivability arguments believe additions can be made that validate the move (Chalmers, 1996), but physicalist critics argue to the contrary (Yablo, 1999; Van Gulick, 1999).

3. *Reduction and emergence can overlap (not necessarily disjoint)*. Although the notions of reduction and emergence are often paired as polar opposites, there are, in fact, many consistent combinations of views from the two respective families. The slogans associated with the two make them seem mutually exclusive. How could Xs 'just be Ys' or 'merely special sorts of Ys' but also be 'something other than Ys'? Or how could Xs be 'something over and above Ys' but also 'nothing more than Ys'? The contradictions seem immediate, and indeed they are if one assumes a consistent reading for both conjuncts. But as we saw above there are many versions of reduction and perhaps equally as many versions of emergence; we distinguished at least ten of each.

Though some versions of reduction are strictly inconsistent with some versions of emergence, other cross combinations involve no necessary conflict. One could not consistently combine an identity version of ontological reduction with a metaphysical notion of radically kind emergent properties or powers. But one might without contradiction accept both a cognitive/explanatory emergence view of mental properties as well as a realization or composition version of ontological reduction.

4. *Nonreductive physicalism lies largely within the intersection of the dual families of emergence and reduction, rather than in a third and separate region.* Nonreductive physicalism is typically regarded as an option wholly distinct from

either reductive physicalism or emergence. However, once one recognizes that reduction and emergence are not mutually exclusive in all their versions, one can see that nonreductive physicalism actually occupies a region within the intersection of the reduction and emergence families. Those two families are to some extent complementaries or duals of each other, especially *vis-à-vis* the subjective/objective division. Thus, nonreductive physicalism on one hand combines a denial of (subjective) representational reduction with an acceptance of some robust form of (objective) ontological reduction such as physical realization. On the other hand, it pairs the denial of at least the strongest forms of ontological emergence such as radical kind emergence with an acceptance of epistemic emergence in either or both of its forms. Put symbolically:

Nonreductive Physicalism => (Ont Reduction & Not REP-Reduction).

Nonreductive Physicalism => (Epistemic Emergence & Not Radical Metaphysical Emergence).

Despite the 'non' in the name of their position, nonreductive physicalists do accept some forms of objective (i.e., ontological) reduction, while rejecting most forms of subjective (i.e., representational) reduction. In the dual domain of emergence, they take the complementary position. They accept various forms of subjective (i.e., epistemic) emergence but reject the radical versions of objective (i.e., metaphysical) emergence as shown in figure 9.

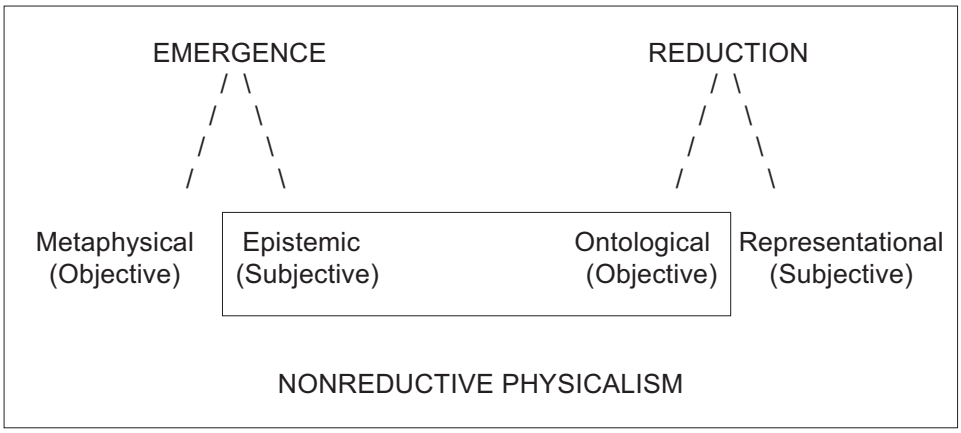

Figure 9

Thus, contrary to common belief, our map of logical space should not locate nonreductive physicalism in a region disjoint and separate from those occupied by the emergence and reduction families, but rather in a special sub-region of their intersection.

5. *What looks like a gap (an anomaly or a crisis) depends on your viewpoint and location.* As noted in the Kuhnian introduction to this paper, much of the pressure for extraordinary philosophical theorizing about consciousness and the psycho–physical link arises from the sense of a persistent anomaly that resists satisfactory resolution by more mainstream physicalist approaches.

There are, no doubt, other factors at work as well. Our current scientific knowledge about the psycho–physical link is admittedly primitive, and there are at present no plausible detailed empirical models of how consciousness might arise from a wholly physical substrate. Research has not advanced far enough to generate a consensus paradigm around which a body of 'normal scientific' practice might coalesce. Extraordinary theorizing is the rule at such early stages, as multiple investigators strike off in differing directions in search of some means to gain significant initial progress on the problem. Such an unsettled scientific state might by itself elicit a good deal of extraordinary philosophic speculation, even if the philosophic community remained largely committed to some form of mainstream physicalism as its own normal practice paradigm. The impulse to engage in extraordinary theorizing is likely contagious and easily spread from one domain to another.

Moreover, extraordinary theorizing need not have any specific trigger at all. Paul Feyerabend, the iconoclastic philosopher of science and champion of revolutionary science, argued that extraordinary science is not and should not be restricted to periods of crisis and anomaly (Feyerabend, 1975). We need not wait until prevailing research programmes 'break' or fail to make progress before exploring radically alternative approaches to the field. Feyerabend had no sympathy with an 'If it's not broke, don't try to fix it' approach to science. He believed that some level of extraordinary or revolutionary science should be going on at all times. Perhaps a good alternative slogan for the Feyerabendian view might be 'Even if it's not broke, you might find something better', or 'If you don't start looking till it is broke, you're not likely to have anything to put in its place when it breaks'. Thus, there are at least two possible sources for extraordinary philosophic theorizing about the psycho–physical link other than the supposed perception of a recalcitrant anomaly. Nonetheless, extraordinary activity is far more likely to occur in contexts of anomaly and perceived crisis, and such a sense does seem to exert a lot of pressure on the current state of philosophic play on the mind/body issue.

But is the perception correct or well founded? Is mainstream physicalism in fact in a crisis caused by its persistent inability to resolve a problem or puzzle that falls clearly within its domain? The answer is not as obvious as it might at first seem. Surely there are puzzles it has not solved and questions to which it has given no satisfying answers, but not every failure to solve a problem counts as an anomaly. As Kuhn (1962) made clear from the start, what counts as a problem or its solution is determined on the whole internally by the research community itself through its paradigms of normal practice. Unless a puzzle presents a problem of a sort that the community by its own internal standards counts as one it ought to be able to solve, its failure to do so need not generate any anomaly or sense of crisis. And even if a puzzle meets the standards to count as a valid problem, it need not be open to a quick or easy solution. If so, a string of negative attempts need not indicate anything more than the difficulty of the problem and the need for continuing effort from within the paradigm.

Physicalism's supporters and its critics will obviously differ in their assessments of its current status. Its critics will argue that it is no closer to explaining

how consciousness might be a physical process than it was three hundred and fifty years ago when Descartes and Leibniz expressed their early scepticism. Nor is it just the explanatory details that we lack. As Thomas Nagel pointed out more then twenty-five years ago in his famous 'What is it like to be a bat?' article (1974), we have a model of how to begin to bring the two sides of the psycho–physical divide together; physicalists still find themselves staring at an explanatory blank wall. Solutions to hard problems rarely come quickly, but if three centuries of failure do not suffice to generate a crisis, what more is needed? Physicalists in reply might compare their inability at the end of the twentieth century to explain the physical nature of consciousness with that of their predecessors at the start of the century to do the same for life. Vitalism was still a serious scientific position one hundred years ago, and a sense of mystery and bafflement still attended attempts to explain life, growth and reproduction as wholly physical. It was only at mid-century that the puzzle was solved when the modern biochemical revolution provided us with more adequate concepts of both the biological processes and their physical substrates. Only when we had a better understanding of both sides of the equation could we see how they fit together. Optimistic physicalists hope that the twenty-first century will do the same for the psycho–physical link. Predictions about what sorts of explanations the future will provide are typically based as much on hope and prior outlook as on evidence, and at present there is no way to decide between optimistic physicalists or their pessimistic critics.

However, it is worth noting that on the Kuhnian model, crises get resolved in many ways, and not all resolutions involve a solution to the problem. They end in at least four common ways:

(1) Through a revolution which solves or rejects the resistant problem.

(2) Through an internal solution of the problem.

(3) Through postponement; putting the problem off for later solution when the field is more advanced.

(4) Through an internal dissolution/rejection of the problem.

How then should we think about the alleged failure to solve the mind/body problem, especially the supposed failure of physicalists to provide any adequate solution to the consciousness/brain version of the problem? There are no doubt major questions as yet unanswered about that relation. At an abstract level, physicalists may claim that consciousness is realized by a wholly physical substrate, but they cannot at this point offer an intuitively satisfying story about how such a realization account might work. Constructing such a model may be more a scientific job than a philosophical one, but in the absence of any such concrete model, abstract physicalist claims about realization ring a little hollow. Perhaps solutions can be found more-or-less within the normal physicalist paradigm, either in the near term or later at a more advanced stage of theorizing, but perhaps not. The solution may instead require a revolution in our ways of thinking about mind, matter, or both. We cannot at this point predict with any great confidence whether the apparent crisis will end in a type 1, 2, or 3 way.

However, there are some explanatory demands that physicalists, at least those accepting some form of mainstream nonreductive physicalism, can justifiably reject. In particular, they should resist the claim that solving the mind/body problem requires deriving our mentalistic theories from our physicalist ones, as reductive logical empiricists had argued. Nor should they agree that doing so must provide us with reductive explanations that allow to see on purely *a priori* conceptual grounds how the physical facts necessitate their conscious consequents, as the dualist proponents of conceivability arguments assume.

Given a pragmatic and contextual account of representation and explanation of the sort that nonreductive physicalists typically hold for quite independent and general reasons, there are good reasons to reject the logical empiricist and apriorist views of what an adequate solution requires. Thus, failures to solve the mind/body problem relative to those representationally reductive criteria need not embarrass the nonreductive physicalist nor count as any sort of anomaly from her perspective. Alleged crises based on our inabilities to provide such reductive accounts might end through dissolution of type 4 rather than a solution; the physicalist may reject the demand for any such account as illegitimate.

Just like other normal practice problem solvers, physicalists get a say in what counts as a problem and a solution within their domain, especially when those views are not *ad hoc* or merely self-protective but are instead based on independently motivated claims about the nature of explanation. Of course, having an internal say in what counts as a solution does not in itself guarantee that outsiders will or should share your evaluation, any more than most of us would be swayed by an astrologer's self-proclaimed success in solving astrological problems in astrologically valid ways. Mainstream physicalists have a responsibility to support and defend their criteria for solving the mind/body problem, but they have plenty of resources for doing so, and their critics have an equal obligation to defend their more reductive standards.

Thus, an accurate map of field should reflect the fact that which problems count as solved, unsolved or illegitimate depends upon one's own location within the space of possible positions. What links one thinks can or cannot be seen will turn in part upon the lie the land appears to have from where one stands. Nor need there be an outside point of view that shows with neutral truth what links there are; there may be only inside answers to the question.

References

Beckermann, A., Flohr, H. and Kim, J. (ed. 1992), *Emergence or Reduction? Essays on the Prospects for Nonreductive Physicalism* (Berlin: DeGruyter).

Block, N. and Stalnaker, R. (1999), 'Conceptual analysis, dualism, and the explanatory gap', *Philosophical Review*, **108**, pp.1–46.

Boyd, R. (1980), 'Materialism with reductionism: what physicalism does not entail', in *Readings in Philosophy of Psychology Volume 1*, ed. N. Block (Cambridge, MA: Harvard University Press).

Chalmers, D. (1996), *The Conscious Mind* (Oxford: Oxford University Press).

Churchland, P.M. (1981), 'Eliminative materialism and the propositional attitudes', *Journal of Philosophy*, **78**, pp. 67–90.

Churchland, P.M. (1985), 'Reduction, qualia, and the direct introspection of brain states, *Journal of Philosophy*, **82**, pp. 8–28.

Davidson, D. (1970), ' Mental events', in *Experience and Theory*, ed. L. Foster and J. Swanson (Amherst: University of Massachusetts Press). Reprinted in Davidson (1980).
Davidson, D. (1974), 'Psychology as philosophy' in *Philosophy of Psychology*, ed. S.C. Brown (New York: MacMiIllan). Reprinted in Davidson (1980).
Davidson D. (1980), *Essays on Actions and Events* (Oxford: Oxford University Press).
Dennett, D. (1988), 'Quining qualia', in Marcel and Bisiach (1988).
Descartes, R. (1642), *Meditations on First Philosophy* (Paris)
Feigl, H. (1958), ' The mental and the physical', in *Minnesota Studies in the Philosophy of Science Volume II*, ed. H. Feigl, M. Scriven and G. Maxwell (Minneapolis: University of Minnesota Press).
Feyerabend, P. (1975), *Against Method* (London: New Left Books).
Fodor, J. (1974), 'Special sciences, or the disunity of science as a working hypothesis', *Synthese* , **28**, pp. 77–115. Reprinted in Fodor (1981).
Fodor, J. (1981), *Representations* (Cambridge, MA: MIT Press).
Foster, J. (1991), *The Immaterial Self* (London: Routledge).
Frege, G. (1892), 'Über Sinn und Bedeutung', *Zeitschrift für Philosophie und Philosophische Kritik*, **100**, pp. 25–50. Translated by Max Black (1952) as 'On sense and reference', in *Translations from the Philosophical Writings of Gottlob Frege*, ed. P. Geach and M. Black (Oxford: Oxford University Press).
Garfinkel, A. (1981), *Forms of Explanation* (New Haven, CT: Yale University Press).
Hasker, W. (1999), *The Emergent Self* (Ithaca: Cornell University Press).
Hill, C. and McLaughlin, B. (1999), 'There is less in reality than dreamt of in Chalmers' philosophy', *Philosophy and Phenomneological Research*, **59**, pp. 445–54
Jackson, F. (1982), 'Epiphenomenal qualia', *Philosophical Quarterly*, **32**, pp.127–36.
Jackson, F. (1986), 'What Mary didn't know', *Journal of Philosophy*, **32**, pp. 291–5.
Jackson, F. (1998), 'Postscript on qualia,' in *Mind, Method and Conditionals,* ed. F. Jackson (London: Routledge).
Kemeny, J. and Oppenheim, P. (1956), 'On reduction', *Philosophical Studies*, **7**, pp. 6–17.
Kim, J. (1982), 'Psychophysical supervenience', *Philosophical Studies*, **41**, pp. 51–70.
Kim, J. (1989), 'The myth of nonreductive physicalism', *Proceedings and Addresses of the American Philosophical Association*, **63**, pp. 31–47. Reprinted in Kim (1993a).
Kim, J. (1990), 'Explanatory exclusion, and the problem of mental causation', in I*nformation, Semantics and Epistemology*, ed. E. Villanueva (Oxford: Basil Blackwell).
Kim, J. (1992), '"Downward causation" and emergence', in *Emergence or Reduction? Essays on the Prospects for Nonreductive Physicalism*, ed. A. Beckermann, H. Flohr and J. Kim (Berlin: DeGruyter).
Kim, J. (1993a), *Supervenience and Mind* (Cambridge: Cambridge University Press).
Kim, J. (1993b), 'Multiple realization and the metaphysics of reduction', in *Supervenience and Mind*, ed. J. Kim (Cambridge: Cambridge University Press).
Kim, J. (1999), *Mind in a Physical World* (Cambridge, MA: MIT Press).
Kripke, S. (1972), *Naming and Necessity* (Cambridge, MA: Harvard University Press).
Kuhn, T. (1962), *The Structure of Scientific Revolutions* (Chicago: University of Chicago Press).
Levin, J. (1986), 'Could love be like a heat wave? Physicalism and the subjective character of experience', *Philosophical Studies*, **49**, pp. 245–61.
Levine, J. (1983), 'Materialism and qualia: the explanatory gap', *Pacific Philosophical Quarterly*, **64**, pp. 354–61.
Levine, J. (1993), 'On leaving out what it's like', in *Consciousness*, ed. M. Davies and G. Humphreys (Oxford: Blackwell).
Lewis, D (1982), 'Postscript to mad pain and Martian pain', in *Philosophical Papers Volume I*, ed. D. Lewis (Oxford: Oxford University Press).
Loar, B. (1990), 'Phenomenal states', in *Philosophical Perspectives Volume 4: Action Theory and the Philosophy of Mind*, ed. J. Tomberlin (Atascadero, CA: Ridgeview Publishing).
Locke, J. (1688), *An Essay on Human Understanding* (London).
Lockwood, M. (1989), *Mind, Brain and Quantum* (Oxford: Blackwell).
Lycan, W. (1990), 'What is the "subjectivity" of the mental?', in *Philosophical Perspectives Volume 4: Action Theory and the Philosophy of Mind*, ed. J. Tomberlin (Atascadero, CA: Ridgeview Publishing).
Marcel, A. and Bisiach E. (1988), *Consciousness in Contemporary Science* (Oxford: Clarendon Press).
McGinn, Colin (1989), 'Can we solve the mind–body problem?', *Mind* , **98**, pp. 349–66.
McGinn, Colin (1991), *The Problem of Consciousness* (Oxford: Blackwell).
McGinn, Colin (1995), 'Consciousness and space', in *Conscious Experience*, ed. T. Metzinger (Thorverton: Imprint Academic).
Moore, G.E. (1902), *Principia Ethica* (Cambridge: Cambridge University Press).
Nagel, E. (1961), *The Structure of Science* (New York: Harcourt, Brace and World).
Nagel, T. (1974), 'What is it like to be a bat?', *Philosophical Review,* Reprinted in *Mortal Questions*,ed. T. Nagel (Cambridge: Cambridge University Press).
Nagel, T. (1979), 'Panpsychism', in *Mortal Questions,* ed. T. Nagel (Cambridge: Cambridge University Press).
Nagel, T. (1986), *The View from Nowhere* (Oxford: Oxford University Press).

Oppenheim, P. and Putnam, H. (1958), 'Unity of science as a working hypothesis', in *Minnesota Studies in the Philosophy of Science Volume II*, ed. H. Feigl, M. Scriven and G. Maxwell (Minneapolis: University of Minnesota Press.
Penrose, R. (1989), *The Emperor's New Mind* (Oxford: Oxford University Press).
Penrose, R. (1994), *Shadows of the Mind* (Oxford: Oxford University Press).
Place, U.T. (1956), 'Is consciousness a brain process?', *British Journal of Psychology,* **47**, pp. 44–50.
Putnam, H. (1972), 'Philosophy and our mental life', in *Mind, Language and Reality, Philosophical Papers Volume 2*, ed. H. Putnam (London: Cambridge University Press).
Putnam, H. (1978), *Meaning and the Moral Sciences* (London: Routledge and Kegan Paul).
Rorty, R. (1970), 'In defense of eliminative materialism', *The Review of Metaphysics*, **24**, pp.112–21.
Russell, B. (1921), *The Analysis of Mind* (London: Macmillan).
Searle, John (1992), *The Rediscovery of the Mind* (Cambridge, MA: MIT Press).
Silberstein, M. (1999), 'Emergence and the mind–body problem', J*ournal of Consciousness Studies*, **5** (4), pp. 464–82.
Silberstein, M. and McGeever, J. (1999), The search for ontological emergence', *The Philosophical Quarterly*, **49**, pp. 182–200.
Smart, J.J.C. (1959), 'Sensations and brain processes', *Philosophical Review*, **82**, pp. 141–56.
Sperry, R. (1983), *Science and Moral Priority: Merging Mind, Brain and Human Values* (New York: Columbia University Press).
Sperry, R. (1991), 'In defense of materialism and emergent interaction', *Journal of Mind and Behavior*, **12**, pp. 221–45.
Strawson, G. (1994), *Mental Reality* (Cambridge, MA: MIT Press).
Swinburne, R. (1986), *The Evolution of the Soul* (Oxford: Oxford University Press).
Van Gulick, R. (1985), 'Physicalism and the subjectivity of the mental', *Philosophical Topics*, **12**, pp. 51–70.
Van Gulick, R. (1992), 'Nonreductive materialism and intertheoretical constraint', in Beckermann *et al.* (1992).
Van Gulick, R. (1993a), 'Understanding the phenomenal mind. Are we all just armadillos?', in *Consciousness*, ed. M. Davies and G. Humphreys (Oxford: Blackwell).
Van Gulick, R. (1993b), 'Who's in charge here and who's doing all the work?', in *Mental Causation*, ed. J. Heil and A. Mele (Oxford: Clarendon Press).
Van Gulick, R. (1999), 'Conceiving beyond our means: the limits of thought experiments', in *Toward a Science of Consciousness III*, ed. S. Hameroff, A. Kazniak and D. Chalmers (Cambridge MA: MIT Press).
Wilkes, K. (1988), '—, yishi, duh, um and consciousness', in Marcel and Bisiach (1988).
Wilkes, K. (1995), 'Losing consciousness', in *Conscious Experience*, ed. T. Metzinger (Thorverton: Imprint Academic).
Yablo, S. (1999), 'Concepts and consciousness', *Philosophy and Phenomenological Research*, **59**, pp. 455–63.