This is the penultimate draft of a chapter for the *Continuum Companion to the Philosophy of Science*, edited by S. French and J. Saatsi

Reduction, Multiple Realizability, and Levels of Reality[*]

Sven Walter and Markus Eronen

The idea of reduction has appeared in different forms throughout the history of science and philosophy. Thales took water to be the fundamental principle of all things; Leucippus and Democritus argued that everything is composed of small, indivisible atoms; Galileo and Newton tried to explain all motion with a few basic laws; 17th century mechanism conceived of everything in terms of the motions and collisions of particles of matter; British Empiricism held that all knowledge is, at root, experiential knowledge; current physicists are searching for the GUT, the "grand unified theory," that will show that at very high energies the electromagnetic and the weak and strong nuclear forces are fused into a single unified field. Some of these projects are clearly ontological in nature (Leucippus and Democritus), others are more methodological (mechanism), and still others strive for theoretical simplification (the projects of Galileo and Newton or the search for a GUT). Nevertheless, as they all aim at revealing some kind of unity or simplicity behind the appearance of plurality or complexity, they may all be regarded as (attempted) reductions.

---

[*] We are grateful to the editors and to Vera Hoffmann, Bob Richardson, and Dan Brooks for helpful comments.

Section 1 surveys philosophical accounts of reduction, focusing mostly on theory reduction, but taking into account ontological aspects of reduction as well. Section 2 addresses the question whether (and if so, how) the special sciences are reducible to more fundamental sciences, in particular in the light of the fact that special science properties seem to be multiply realizable. Section 3 looks at some attempts to understand reductive endeavors in terms of mechanistic explanations. Section 4 explores the interconnections between scientific reductions and the idea that our world is a layered one with distinguishable levels of organization. Section 5 finally briefly highlights some worthwhile future research questions.

Space limitations prevent us from addressing some issues pertinent to the topic, like emergence (Bedau & Humphreys 2008; Stephan 2006), properties and powers (Molnar 2003) or laws (Carroll 1994; Mumford 2004). Although we will mostly be concerned with discussions in the philosophy of mind having to do with the reduction of psychology to neuroscience, the issues we raise are rather general and arise in other disciplines as well, including macrophysics (Batterman 2000, 2001), biology (Brigandt forthcoming; Brigandt & Love 2008; Schaffner 1993; Wimsatt 2007), chemistry (Hendry 2009), and the social sciences (Jackson & Pettit 1992).

1. Theory Reduction

1.1 Reduction as Translation

In the early 20th century, logical positivists set out to understand the nature of science and the relations between the various sciences. One of their goals was to "unify science" by finding a common language into which all meaningful scientific statements are translatable. Rudolf Carnap (1932*a*, 1932*b*) and Carl Gustav Hempel (1949) argued that the language of physics could serve as the universal language of science. Meaningful scientific concepts, statements, and laws, they held, must be translatable into physical concepts, statements, and laws. Psychology, for example, is “an integral part of physics” in that “[a]ll psychological statements which are meaningful … are translatable into statements which do not involve psychological concepts, but only the concepts of physics” (Hempel 1949, 18). The psychological predicate “*x* is excited,” for example, Carnap (1932*b*, 170–171) argued, is translatable into a physical predicate like “*x*’s body … has a physical structure that is characterized by a high pulse and rate of breathing, by vehement and factually unsatisfactory answers to questions, by the occurrence of agitated movements on the application of certain stimuli etc.” Carnap defended this claim by pointing out, first, that verificationism entails that predicates are synonymous iff they are applied on the basis of the same observations and, secondly, that the physical predicate simply enumerates the observations on the basis of which the psychological is applied. Since Carnap took properties to be the intensions of predicates, he thought that properties were identical iff the corresponding predicates were synonymous and thus that the translatability claim also vindicated an *ontological reduction* of the property of being excited.

The hope that physics could serve as a *lingua franca* of science was soon dashed, however, because many *prima vista* meaningful statements of the special sciences, including those of psychology, were simply not translatable into physical language non-circularly: notoriously, someone who *wants* a beer will go to the fridge to get one only if she believes that there is beer in the fridge, does not attempt to stay sober etc. Synonymy thus seemed to be too strong a requirement on both the theoretical reduction of psychology to physics and the ontological reduction of the mental to the physical. In the philosophy of science translational reduction was therefore replaced by more sophisticated models of reduction (see below), while the philosophy of mind abandoned synonymy as a prerequisite for property-identities, paving the way for the idea that psychophysical property-identities are what Kripke (1980) called "a posteriori necessities" (see section 2).

1.2 Oppenheim and Putnam: The Unity of Science

Although the dream of a wholesale translation of all scientific statements into the language of physics had to be given up, the ideal of a unified science in which special sciences like chemistry, biology, psychology etc. are reducible to more fundamental theories was retained. Could something like a "unity of science" not be attained even if higher-level predicates like "x is soluble," "x is a Chinese wisteria," or "x is excited" are not translatable into purely physical terminology? Paul Oppenheim and Hilary Putnam (1958) suggested as a "working hypothesis" the view that all sciences are reducible to physics via a series of microreductions. Theory T_2 microreduces to theory T_1 iff (1.) any observational data explainable by T_2 are explainable by T_1, (2.) T_1 has more "systematic power" than T_2, and (3.) all the entities referred to in T_2 are fully decomposable into entities belonging to the universe of discourse of T_1. Oppenheim and Putnam's approach faced severe difficulties. For example, it is unclear whether the observational and the non-observational can always be clearly distinguished, the notion of "systematic power" is not clearly defined, and there are hardly any historical cases that satisfy the proposed conditions (Sklar 1967). Nevertheless, many of its key ideas are still visible in what became the standard model of intertheoretic reduction for decades to come: Ernest Nagel's (1961) model of reductions as derivations via bridge-laws.

1.3 Nagel: Reductions as Derivations via Bridge-laws

Nagel (especially 1961, 336–397) took the idea that reduction consists in the derivation of the reduced theory T_2 from a reducing theory T_1 seriously. Such derivations are possible, Nagel (1961, 352–356) argued, if (1.) the terms of T_2 are connectable with terms of T_1 by means of suitable bridge-laws, i.e. empirical hypotheses that express material rather than logical connections (the "condition of connectability"), and (2.) given these connecting principles, all laws of T_2 can be derived from laws of T_1 (the "condition of derivability"). Reductions could thus be seen as deductive-nomological explanations, where T_1 explains T_2. Since in all interesting cases of reduction, T_1 and T_2 are going to be framed in partially disjoint vocabularies, the connectability condition is essential: without connecting bridge-laws the required derivations would be impossible. The exact nature of bridge-laws has been a matter of debate. Although Nagel allowed them to be material conditionals of the form "$\forall x\ (F_{T1}x \supset F_{T2}x)$" (Richardson 1979), it was usually assumed that biconditionals of the form "$\forall x\ (F_{T1}x \equiv F_{T2}x)$" are necessary for the ontological simplifications that were considered to be one of the main goals of reduction (see section 2).

Unlike Oppenheim and Putnam's approach, Nagel's model of reduction was formally precise, but it also failed to fit standard cases of scientific reduction. Nagel himself acknowledged that even in his own example, the reduction of thermodynamics to statistical mechanics, the actual derivation would be immensely complicated and possible only under a set of idealizing assumptions (one has to assume, e.g., that the gas is composed of a large number of perfectly elastic spherical molecules with equal masses and volumes that are in constant motion and subject only to forces of impact between themselves and the walls of the container). In fact, the derivation may not be possible at all (Richardson 2007; Sklar 1999), given that central thermodynamical concepts like "entropy" are associated with a variety of distinct concepts in statistical mechanics which do not exactly correspond to "entropy," neither separately nor taken together.

Another important problem for Nagel's account was that the reducing theory often <u>corrects</u> the original theory, which entails that the original theory was <u>false</u>. For example, Newtonian physics showed that some principles of Galilean physics, like the assumption that uniformly accelerated gravitational free-fall is the fundamental law of motion, were false. However, since logical deduction is truth-preserving, the new, reducing, theory cannot both be true and logically entail a false theory. Problems like these led Paul Feyerabend (1962) to argue that no formal accounts of scientific reduction are possible or necessary. The majority of philosophers, however, responded by developing more sophisticated models (Causey 1977; Schaffner 1967), culminating in what became known as, using John Bickle's (1998) term, "New Wave Reductionism" (NWR).

1.4 New Wave Reductionism

Like its precursor, NWR is an allegedly universal model that takes reduction to be a relation involving logical derivations between theories (Bickle 1998, 2003; Hooker 1981; see also Churchland 1985; Churchland 1986; Schaffner 1993). However, what is derived from T_1 is not T_2 itself, but an "equipotent isomorphic image" T_2a of T_2, which renders the falsity of T_2 (see section 1.3) unproblematic. The ultimate fate of T_2 and its ontological posits depends upon the exact relation between T_2 and T_2a. If the analogy between T_2 and T_2a is strong, not much correction is needed. In that case, T_2 is reduced "smoothly" to T_1 and T_2a retains many of the entities posited by T_2. In contrast, if T_2 and T_2a are only weakly analogous, the amount of correction needed is considerable. In that case, the reduction is "bumpy" and many or all of the entities posited by T_2 will be eliminated from the ontology of T_2a. It is not clear, however, how exactly to evaluate the strength of the analogy between T_2 and T_2a. Additionally, NWR inherits two problems that already plagued early approaches.

First, NWR is still intended as a general model of scientific reduction. This renders it blind to certain fundamental differences (McCauley 2007; Nickles 1973; Wimsatt 1976). Most importantly, it fails to account for the difference between intralevel (or successional) relations between competing theories within a particular science (e.g., Newtonian theory of gravity and general relativity theory) on the one and interlevel relations between theories (e.g., cognitive psychology and cellular neuroscience) on the other hand. In particular, the examples of eliminative, or "bumpy," reductions offered by NWR are all intralevel cases and thus provide no reason to expect eliminative reductions in interlevel contexts, for example between psychology and neuroscience.

Second, NWR retains the idea that the relata of reductions are formal or at least semiformal theories, phrased in first-order predicate logic or set-theoretic terms. Yet, some generally accepted cases of scientific reduction—the reduction of genetics to molecular biology, say—do not seem to involve such formal theories (Sarkar 1992). Quite generally, while the formal theories that are suitable as starting points of logical derivations may be available in theoretical physics, most special sciences simply do not have any well-structured theories that could be handled formally (see, however, Schaffner (1993) for a defense of formal approaches to reduction in biology). Explanations and reductions in these disciplines can hardly be conceived as logical derivations. Instead, these disciplines typically look for descriptions of mechanisms that can serve as explanations for patterns, effects, capacities or phenomena (section 3).

2. Multiple Realizability and the Reduction of Special Sciences

2.1 Multiple Realizability and Kim's Dilemma

In the philosophy of mind, the issue of reduction surfaces in the debate between reductionists and non-reductionists. While reductionists hold that the mental can be reduced to the physical—at least ontologically, if not conceptually—non-reductionists maintain that although such reductions fail, mental properties are nevertheless not non-physical in any ontologically threatening sense: the mental is irreducible, and thus ontologically and conceptually autonomous, but since it is realized by, dependent upon, or supervenient upon the physical, it is "naturalistically kosher."

Once psychophysical predicate synonymies turned out be unattainable (see section 1), early identity theorists famously argued for *a posteriori* identities. The thesis that consciousness is a brain process, Place (1956, 45) held, is not a consequence of a successful conceptual reduction in which mentalistic statements are shown to follow *a priori* from statements couched in physical terms only, but a "reasonable scientific hypothesis," on a par with other theoretical identifications *a posteriori* like "Water is H_2O." Putnam (1967) objected that the *a posteriori* identity of mental and physical properties is an ambitious and probably false hypothesis because mental properties are *multiply realizable* by different physical properties in different species, conspecifics, and even one individual at different times. Fodor (1974) provided further support for non-reductionism, arguing that Putnam's considerations apply to all special science properties. According to what Fodor (1974, 97) called the "generality of physics," all entities subsumed under special science laws must at root be *physical* entities. Yet, since a special science property M will typically be multiply realizable, statements like "$(\forall \underline{x})(M\underline{x} \equiv P\underline{x})$" linking M with a physical property P will usually be false, and hence fail to be laws. Statements like "$(\forall \underline{x})(M\underline{x} \equiv (P_1\underline{x} \vee \dots \vee P_n\underline{x})$" linking M with the complete disjunction of all of its physical realizers will be true, but cannot be laws either, because "$(P_1\underline{x} \vee \dots \vee P_n\underline{x})$" fails to designate a scientific kind (see section 2.2). Hence, there are no laws—and thus *a fortiori* no *bridge-laws*—connecting special science properties with physical kinds. This renders Nagelian reductions of special science properties impossible.

Due mostly to the arguments of Putnam and Fodor, non-reductionism achieved an almost hegemonic status during the 1970s and 1980s. Jaegwon Kim (1992), however, forcefully argued that far from making reductions impossible, multiple realizability actually engenders them. Non-reductionists, he maintained, face the following dilemma.

On the one hand, if Fodor is wrong and a disjunctive predicate like “$P_1x \vee \ldots \vee P_nx$” designates a kind, then nothing prevents us from reducing a multiply realizable special science property *via* a disjunctive bridge-law. Call this the “*Disjunctive Move*.” On the other hand, if Fodor is right and “$P_1x \vee \ldots \vee P_nx$” does *not* designate a scientific kind, then the predicate “M” with which it is coextensive cannot designate a scientific kind either. If there are to be any special science laws at all, they must thus be couched in terms of the only law-fit predicates left, viz., “P_1,” “P_2,” …, “P_n.” This leads to so-called “local,” or “species-specific,” reductions *via* bridge-laws of the form “$(\forall x)\ (Sx \supset (Mx \equiv Px))$” saying that if *x* belongs to species S, then *x* has M iff *x* has P. Call this “*Local Reductionism*” (if M is multiply realizable below the level of species, “S” refers to individuals, individuals at times etc.). On either horn, Kim argued, reductionism carries the day.

Another important attack on non-reductionism has come from authors who argue that multiple realizability is in fact not at all common in the special sciences. William Bechtel and Jennifer Mundale (1999, 176–177), for example, claimed that “a proper examination of neurobiological and cognitive neuroscience practice will show that the claim that psychological states are in fact multiply realized is unjustified, and that what is usually taken to be evidence for it, is not.” In a similar vein, Bickle (2003, ch. 3, esp. 131–158) argued that the cellular mechanisms underlying memory consolidation are the same in fruit flies, sea slugs and rabbits, and Batterman (2000) has made a similar point with regard to the alleged multiple realizability of macrophysical properties (for a philosophical defense of the claim that the thesis of multiple realizability has been oversold see Shapiro (2004)).

2.2 The Disjunctive Move

Although the inadequacy of bridge-law based approaches to reduction was evident in philosophy of science and Nagel's account had already been replaced by more sophisticated models, the debate between reductionism and non-reductionism in the philosophy of mind was concerned with the availability of bridge-laws until the late 1990s before finally alternative models of reduction were explored (see section 2.4).

Proponents of the Disjunctive Move, for example, assumed with Nagel that the existence of bridge-laws linking mental and physical predicates is sufficient for reductions, and then argued that multiple realizability is compatible with reductions because there will always be true biconditionals linking mental properties with the complete disjunction of their physical realizers. In response, opponents of the Disjunctive Move tried to show that such biconditionals cannot be bridge-laws.

According to a traditional (though not universally accepted) view of laws, they exhibit two features that have been said to cause trouble for the Disjunctive Move. "$(\forall \underline{x})(F\underline{x} \equiv G\underline{x})$" is a law only if (1.) it is explanatory, and (2.) "F" and "G" are projectible in the sense that observations of Fs which are G increase confidence that the next observed F will also be G. Opponents of the Disjunctive Move have argued that disjunctive "laws" fail on both counts. "$(\forall \underline{x})(M\underline{x} \equiv (P_1\underline{x} \vee \ldots \vee P_n\underline{x}))$" is not explanatory (Pereboom & Kornblith 1991; but see Jaworski 2002), and the predicate "$P_1\underline{x} \vee \ldots \vee P_n\underline{x}$" is unprojectible because is causally heterogeneous: from a causal point of view, there is nothing in common to all and only the individuals satisfying it (Fodor 1974; Kim 1992; 1998, 106–110; but see Walter 2006).

Given this, the prospects for the Disjunctive Move seem dim. In terms of Kim's dilemma, however, this only leads to the second horn, viz., to Local Reductionism.

2.3 Local Reductionism

According to Kim, if the disjunction of M's physical realizers is causally heterogeneous, unprojectible, and thus non-nomic, then M (say, the property having pain) cannot be a nomic property either, given that these properties are instantiated by the same individuals in all nomologically possible worlds: "If pain is nomically equivalent to [a] property claimed to be wildly disjunctive and obviously non-nomic, why isn't pain itself equally heterogeneous and non-nomic as a kind? … It is difficult to see how one could have it both ways—i.e., to castigate [the latter] as unacceptably disjunctive while insisting on the integrity of pain as a scientific kind" (Kim 1992, 323–324). This insight, Kim argued, leads to a positive account of reduction. Consider P_h, P_r and P_m, the physical realizers of having pain in humans, reptiles, and Martians. Suppose P_h, P_r and P_m considered individually are causally homogeneous and thus projectible, but so different that the disjunction $P_h \vee P_r \vee P_m$ is causally heterogeneous, and thus unprojectible and non-nomic. Given Kim's argument that having pain cannot be nomic if $P_h \vee P_r \vee P_m$ is non-nomic, there can thus be no laws about pain as such. The only projectible pain-properties left are P_h, P_r, and P_m, and so the only genuine laws about pain are laws about pain-in-humans, pain-in-reptiles and pain-in-Martians. Hence, "there will be no unified, integrated theory encompassing all pains in all pain-capable organisms, only a conjunction of pain theories for appropriately individuated biological species and physical structure types” (Kim 1992, 325). The result are restricted bridge-laws “$(\forall x)(S_h x \supset (Mx \equiv P_h x))$,” “$(\forall x)(S_r x \supset (Mx \equiv P_r x))$,” and “$(\forall x)(S_m x \supset (Mx \equiv P_m x))$” which sunder the psychological theory about pain in three different subfields, each of which is “locally reducible” (Kim 1992, 328). The same holds *mutadis mutandis* for all other multiply realizable special science properties.

Kim eventually came to reject *Local Reductionism*, however. A successful reduction of *x* to *y*, he held (1998, 96), should be *explanatory* by making intelligible how *x* can arise out of *y* and *simplify ontology* by getting rid of *x* as an entity in its own right. Bridge-laws, however, universal or restricted, fail on both counts. First, even if “$(\forall x)$ (x has pain $\equiv$ x has c-fiber firing)” were a law, this would not explain why having c-fiber firing feels painish rather than ticklish (Kim 1998, 95–96). Second, bridge-laws do not simplify ontology. One reason is that even if “$(\forall x)$ (x has pain $\equiv$ x has c-fiber firing)” were a law, *having pain* and *having c-fiber firing* would still not be identical because the law would be contingent and its contingency could arguably not be blamed on a contingency involving an epistemic counterpart, as in all other cases of scientific identifications (Kripke 1980). Another reason is that even if restricted bridge-laws like “$(\forall x)(S_h x \supset (Mx \equiv P_h x))$” are true and “M” is coextensive with “P_h” relative to S_h and with “P_r” relative to S_r etc., it seems that the *property* M cannot be identical with P_h relative to S_h and with P_r relative to S_r: M is in this context typically construed as a functional property—it is the second-order property of having some first-order property (P_h, P_r etc.) that occupies a certain causal role. But then P_h, P_r etc. and M cannot be identical, for first-order occupants of causal roles cannot be identical to the second-order properties whose causal role they occupy. Therefore, “Nagel reduction gives us no ontological simplification, and fails to give meaning to the intuitive ‘nothing over and above’ that we rightly associate with the idea of reduction” (Kim 1998, 97).

Kim thus became convinced that only genuine property-identities can yield reductions. He therefore modified his Local Reductionism in a way that preserved the key idea that multiple realizability leads to species-specific reductions, while at the same time allowing for genuine species-relative property-identities. The result was his model of Functional Reduction.

2.4 Functional Reduction

Kim’s model of Functional Reduction is based on ideas from David Lewis (1980). Lewis argued that instead of looking for property-identities across all possible worlds, we should identify mental properties with physical properties relative to worlds, species, or structures. The concept “pain,” he claimed, is a functional concept in the filler-functionalism sense, not in the more popular role-functionalism sense. “Pain” is the concept of a property that occupies a causal role, not the concept of the property of having a property that occupies a causal role. In contrast to the usual role-functionalism reading, a filler-functionalism reading of “pain” leads to property identities: “If the concept of pain is the concept of a state that occupies a certain causal role, then whatever state does occupy that role is pain” (Lewis 1980, 218). According to Lewis, “pain” is a non-rigid designator, defined relationally in terms of the causal role of pain, which picks out different physical properties relative to different species. The gerund “being in pain,” in contrast, is the role-functionalist predicate that picks out the same property, viz., the functional property of having a property that occupies the pain-role, in each world, species, or structure (Lewis 1994, 420). Thus, according to Lewis' (not at all uncontested) view, "being in pain" rigidly designates the same functional property in all creatures, whereas "pain" non-rigidly designates different physical fillers of the pain-role in different species.

If "M" means "the occupant of the M-role" and if there is variation in what occupies the M-role, Lewis argued, then not only the contingent laws relating "M" to physical predicates but the property-identities themselves are restricted: "not plain M = P, but M-in-K = P, where K is a kind within which P occupies the M-role. Human pain might be one thing, Martian pain might be something else" (Lewis 1994, 420). Since these are genuine property-identities, Lewis' account yields the ontologically simplifying and explanatory reductions Kim was looking for. Since M-in-K is identical to P, there is no need to recognize M-in-K as a property in its own right, and if P is the property that plays the M-role, there is no question of explaining why M-in-K is correlated with P—having M-in-K just is having P.

Lewis-style reductions are essentially three-step procedures: A special science property M is first construed via conceptual analysis as the property characterized by a certain causal role; then the physical property P occupying that causal role in a world, species, or structure S is identified by means of empirical investigation, and finally M and P are contingently identified, resulting in an identification of M-in-S with P. This became the key idea behind Kim's model of Functional Reduction:

> For functional reduction we construe *M* as a second-order property defined by its causal role … So *M* is now the property of having a property with such-and-such causal potentials, and it turns out that property *P* is exactly the property that fits the causal specification. And this grounds the identification of *M* with *P*. *M* is the property of having some property that meets specification *H*, and *P* is the property that meets *H*. So *M* is the property of having *P*. But in general the property of having property *Q* = property *Q*. It follows then that *M* is *P*. (Kim 1998, 98–99)

Kim's new model allegedly avoids the two problems that prevented bridge-laws from yielding genuine property-identities (see section 2.3): (1.) Kripke's argument concerning the necessity of identities, and (2.) the fact that first- and second-order properties cannot be identical. Kripke's argument is ineffective because it works only for identity statements containing rigid designators, while "*x* has pain" is supposedly *non-rigid*. Furthermore, instead of talking about second-order *properties*, it would be more appropriate to talk about second-order *designators* or *predicates*. Second-order designators express role-concepts that are filled by first-order physical properties, and they (non-rigidly) designate these first-order physical properties, rather than a second-order property common to all individuals that satisfy them. The predicate "*x* has pain" thus expresses the concept "pain," but it neither denotes the property *having pain* nor any other property common to all and only the individuals that have pain (here Kim disagrees with Lewis who acknowledged such a property, viz., the role-functional property expressed by the gerund "being in pain").

Kim's model of *Functional Reduction* is a kind of *eliminative reduction* (see section 1.4). *Having pain* is abandoned as a genuine property which can be exemplified by creatures of different species; there only remain the predicate "x has pain" and the concept "pain" which equivocally pick out distinct properties in different species. Although mental predicates and concepts group physical properties in ways essential for descriptive, explanatory and communicative purposes, we have to learn to live without universal mental properties like *having pain* (Kim 1998, 106). It is thus clear why Kim thought the multiple realizability argument for non-reductionism fails: the differences among the physical realizers of special science properties do not show that these properties are multiply realizable, but that the corresponding predicate non-rigidly picks out more than one property.

One important problem with the model of *Functional Reduction* is that mental properties might be multiply realizable not only in different species, but also in conspecifics or even single individuals so that *having pain* would be one physical property in Paul and another in Peter, or one in Paul at t_1 and another in Paul at t_2. But further narrowing mental kinds into ever more restricted physical structures seems theoretically self-defeating, as with the increasing loss of generality the identifications will be theoretically uninteresting and purely *ad hoc*.

Another problem is that properties which cannot be construed relationally in terms of their causal role—in particular phenomenal properties like *having a reddish visual experience*, *having a lemonish gustatory experience* etc.—will not be susceptible to functional reductions. The problem, however, is not only that such properties turn out to be irreducible, but that Kim's own *Supervenience Argument* (Kim 1998, 2005; see also Walter 2008) is designed to show that irreducible properties cannot be causally efficacious. Kim (2005, 173) has thus reluctantly admitted that phenomenal properties are causally otiose epiphenomena, so that the "fact that blue looks just this way to me, green looks that way, and so on, should make no difference to the primary cognitive function of my visual system."

Finally, Kim presents his model as a realistic general account of reduction in science (Kim 1998, 99), but does not show that scientific cases of reduction actually conform to it. Rueger (2006), for example, argues that Kim's model is inapplicable in physics. Kim's own favorite example is a biological one—the reduction of the property of being a gene to strands of DNA. But even this example is presented only very schematically and in a way that does no justice to actual history and to the philosophy of biology (Hull 1972; Schaffner 1969; Wimsatt 2007, ch. 11). This illustrates once again that discussions about reduction in the philosophy of mind have been largely unconstrained by, and are effectively lagging behind, developments in the philosophy of science.

3. Mechanistic Explanation, Explanatory Pluralism, and Ruthless Reductionism

Mostly due to the reasons outlined in section 1, theory reduction is nowadays not considered to be the norm in the special sciences. What has become something like the new received view on the nature of interlevel and intertheoretic relations is rather what is known as "mechanistic explanation" (Bechtel 2008; Bechtel & Richardson 1993; Craver 2007; Machamer et al. 2000). The basic insight of this approach has already been noted at the end of section 1: if one takes into account actual scientific practice in neuroscience and many of the life sciences, it turns out that instead of focusing on formalizable theories and their derivability from more fundamental ones, practicing scientists try to formulate explanations in terms of empirically discoverable *mechanisms*. Broadly speaking, mechanisms are "entities and activities organized such that they are productive of regular changes from start or set-up to finish or termination conditions" (Machamer *et al*. 2000, 3). Or, as Bechtel (2008, 13) puts it, a "mechanism is a structure performing a function in virtue of its component parts, component operations, and their organization." A mechanistic explanation then describes how the orchestrated functioning of the mechanism is responsible for the phenomenon to be explained.

Consider the example of memory consolidation (Bickle 2003; Craver 2002, 2007). A mechanistic explanation of memory consolidation describes the cellular and molecular mechanisms underlying it by showing how the relevant parts of the memory system and their activities together result in the transformation of short-term into long-term memories. Central to this explanation is *Long Term Potentiation* (LTP), a well-studied cellular and molecular phenomenon that exhibits features that make it very likely the central part of the memory consolidation mechanism.

Typically, mechanistic explanations have to be *multilevel*, because focusing on a single level does not allow for a full understanding of the explanandum. In the case of memory consolidation, for instance, Craver (2002) identifies four relevant levels which, crucially, are not to be understood as general levels of organization, but simply as the levels of the mechanism in question (see section 4): (1.) the *behavioral-organismic level* (involving various types of memory and learning, the conditions for memory consolidation and retrieval etc.); (2.) the computational-hippocampal level (involving structural features of the hippocampus, its connections to other brain regions, and the computational processes it supposedly performs etc.); (3.) the electrical-synaptic level (involving neurons, synapses, dendritic spines, axons, action potentials etc.); and (4.) the molecular-kinetic level (involving glutamate, NMDA and AMPA receptors, Ca^{2+} ions, and Mg^{2+} ions etc.).

Mechanistic explanations have both a "downward-looking" and an "upward-looking" aspect. In the LTP case, one is looking upward when, in order to understand the computational properties of the hippocampus, one is taking into account its environment, or when, in order to understand the role of the molecular processes of LTP, one is looking at the larger computational-hippocampal framework. In contrast, one is looking downward when memory consolidation is explained by appeal to the computational processes at the hippocampal level, or when the synaptic LTP mechanism is explained by appeal to activities at the molecular-kinetic level.

On the one hand, since mechanistic explanation does not necessarily accede primacy to lower levels, it can be seen as supporting a kind of anti-reductionist explanatory pluralism (Craver 2007; McCauley 2007; Richardson & Stephan 2007). This anti-reductionist conclusion receives further support from the "interventionist" account of causation (Woodward 2003, 2008), according to which higher-level entities can have causal and explanatory relevance even if lower-level explanations in terms of implementing mechanisms are complete.

On the other hand, the process of "looking downward" and invoking parts of the mechanism to understand its behavior as a whole is close enough to what scientists generally take to be a reductive explanation to warrant treating the downward-looking aspect of mechanistic explanation as a kind of reductive explanation (Bechtel 2008; Sarkar 1992; Wimsatt 1976). Carl Gillett (2007), for example, argued that mechanistic explanations in fact imply ontological reductions. Also, John Bickle (2003, 2006) has taken the reductive aspect of mechanistic explanations seriously, arguing for what he calls a "ruthlessly reductive" analysis of explanation in neuroscience. According to Bickle, when we look at experimental practices in molecular and cellular cognition, we find a two-step strategy: the researcher (1.) causally intervenes into cellular or molecular pathways in order to (2.) track statistically significant differences in behavior resulting from these interventions. If successful, this strategy establishes a scientific reduction by forging a mind-to-molecules linkage. Importantly, once the lower-level explanations are completed, higher-level sciences are retained only for heuristic and pragmatic purposes: "psychological explanations lose their initial status as causally-mechanistically explanatory vis-à-vis an accomplished ... cellular/molecular explanation" (Bickle 2003, 110). One problem for Bickle's account is that while advocates of explanatory pluralism can appeal to the interventionist account of causation, it is unclear which account of causation or causal explanation Bickle could appeal to. Mechanistic explanation pluralistically understood thus seems to have a stronger case, so that explanation in neuroscience seems, if at all, only "somewhat" reductive—not "ruthlessly reductive," and not eliminative.

4. Reduction and Levels of Reality

Talk of levels is ubiquitous. Philosophers talk about levels of nature, analysis, realization, being, organization, explanation, or existence, to name just a few. In science, the list is even longer. In the neurosciences alone, at least the following uses of the term “level” can be found: levels of abstraction, analysis, behavior, complexity, description, explanation, function, generality, organization, science and theory (Craver 2007, 163–164).

Talk of levels has of course also been important in debates about reduction. Early on (see section 1), when the goal was to reduce all “higher-level” sciences to “lower-level” sciences, one important question was how to sort the various sciences into levels. Oppenheim & Putnam (1958) proposed a preliminary division into six hierarchical levels—social groups, (multicellular) living things, cells, molecules, atoms, and elementary particles—which were supposedly related *mereologically* in the sense that the entities at any given level are composed of entities at the next lower level.

A similar appeal to mereology can nowadays be found in Kim’s work with regard to *levels of properties*. The level of a property, Kim (1998, 92) argued, depends upon what it is a property of: properties of objects with parts are higher-level with regard to the properties of their parts, and properties of objects with no parts are fundamental properties. In addition to that, every level of reality has different *orders* of properties, generated by the supervenience relation: second-order properties are generated by quantification over the first-order properties that form their supervenience base (Kim 1998, 20). Each level thus contains lower- and higher-order properties; higher-order properties are properties supervening upon lower-order properties of the same level, not upon lower-level properties. Supervenience thus generates an *intralevel* hierarchy of lower- and higher-order properties, while the interlevel micro/macro hierarchy between properties of wholes and properties of their parts is not generated by supervenience, but by mereology.

The mereological appeal to composition can be found in nearly all philosophical accounts of levels of organization. In addition, size or scale are often presented as criteria (Churchland & Sejnowski 1992), where organization by size is obviously related to compositional criteria, as parts are smaller or at least no bigger than wholes. However, these criteria lead to anomalies and unwanted conclusions. A pile of snow, for example, is composed of smaller piles, but this does not mean that the larger pile is at a higher level than the smaller ones. Regarding size, there are bacterium-sized black holes and raindrop-sized computers, but it does not seem natural to say that bacteria are at the same level as black holes, or that raindrops are at the same level as tiny computers.

Perhaps the most comprehensive account of levels of organization has been developed by William Wimsatt (1976, 2007). Wimsatt's starting point is that levels of organization are compositional levels that are non-arbitrary features of the ontological architecture of the world. Wimsatt is not aiming at a strict definition of levels, but rather at establishing sort of a "prototype" idea of levels, by characterizing several characteristics levels typically (but not necessarily) have. For example, levels of organization are constituted by families of entities usually of comparable size, and the things at a level mostly interact with other things at the same level, so that the regularities of the behavior of a thing are most economically expressed in terms of variables and properties appropriate for that level. As a kind of a preliminary definition, Wimsatt (2007, 209) suggests that "levels of organization can be thought of as local maxima of *regularity and predictability in the phase space of alternative modes of organization of matter*." Roughly speaking, this means that at the scale of atoms, for example, there are more regularities than at scales just slightly larger or smaller, so that at the scale of atoms there is a peak of regularity and predictability, and thus a level of organization.

However, Wimsatt acknowledges that instead of a neat hierarchy of the Oppenheim & Putnam (1958) kind, these criteria yield a complex and branching structure of levels. Furthermore, at higher levels, for example in psychology and neuroscience, neat compositional relations break down. According to Wimsatt (2007, 227–237), levels become less useful here for characterizing the organization of systems, and it becomes more accurate to talk of "perspectives." Perspectives are subjective or at least quasi-subjective views of systems and their structures that do not give a complete description of all aspects of the systems in question, and that do not map compositionally onto one another as levels of organization do. When even the boundaries of perspectives begin to break down, perspectives degenerate into so called "*causal thickets*" where things are so intertwined and multiply-connected that it is impossible to determine what is composed of what and which perspective a problem belongs to (Wimsatt 2007, 237–240). According to Wimsatt, the neurophysiological, the psychological and social realms are for the most part such causal thickets. Unfortunately, the notions of perspectives and causal thickets remain rather vague and unclear in Wimsatt's account.

Levels also play a central role in the context of mechanistic explanations (see section 3). The levels of mechanistic explanations are a special variety of levels of composition whose relata are mechanisms at higher levels and their components at lower levels (Craver 2007, ch. 5). This notion of level is in one respect fundamentally different from general levels of organization. Levels of mechanisms are not universal divisions in the structure of the world (à la Oppenheim & Putnam). Rather, different mechanisms have different hierarchies of levels. The levels in the spatial memory system, for example, are different from those in the circulatory system. According to the mechanist, these local and case-specific levels are sufficient for understanding reductive explanations and interlevel relations in many fields (Bechtel 2008; Craver 2007). One limitation of this is that global comparisons become impossible. We cannot say that cells are, in general, at a higher level than molecules. All we can say is that cells in a certain mechanism are at a higher level than the molecules that are part of the same mechanism. We cannot even say that a certain molecule in a certain brain is at a lower level than the hippocampus of that brain, unless the molecule is involved in the same mechanism as the hippocampus. Even within a certain mechanism it is not possible to say whether subcomponents of two different components are at the same level or not, since they do not stand in a part-whole relation to each other.

Wimsatt-style levels of organization and levels of mechanisms are not necessarily incompatible. As seen above, levels of organization are said to “break down” in the neurophysiological and the psychological realms, and these are exactly the realms where levels of mechanisms are typically applied. In this sense, the two accounts may simply complement each other.

5. Directions for Future Research

There is a huge gap between formal approaches to reduction (e.g., NWR, <u>Functional Reduction</u>) and non-formal approaches that are closer to scientific practice (e.g., mechanistic explanations). What is still unclear is whether non-formal approaches are able to replace formal analyses entirely or whether they just have to be seen as complementing them. Answering that question convincingly requires a clearer picture of the possible fields of application of non-formal and formal approaches than we as of yet have, and a better understanding of the limitations, theoretically and practically, of the non-formal approaches.

Another cluster of important open questions concerns the idea of levels, a topic that has received comparatively little philosophical attention. Are there universal levels of organization or just local, case-specific, levels? What are the criteria for assigning things to levels? In what sense, if any, are lower-level explanations and theories more fundamental than higher-level ones? These questions in turn are related to some well-known but still unresolved metaphysical debates in the philosophy of mind that could benefit enormously from finally taking into account actual scientific practice: Is there any multiple realizability, and if so, what is its import? What are the ontological implications of successful mechanistic explanations? Is explanatory pluralism and higher-level (interventionist) causation compatible with physicalism? In particular, what are its implications with regard to the causal closure of the micro-physical and the denial of overdetermination?

Some of these questions are purely philosophical, others have clear empirical aspects, but they are all of prime importance to the philosophy of science and should be targets of further research.

References

Batterman, R. (2000). Multiple realizability and universality. *British Journal for the Philosophy of Science* 51, 115–145.

Batterman, R. (2001). *The Devil in the Details*. Oxford: Oxford University Press.

Bechtel, W. (2008). *Mental Mechanisms*. London: Routledge.

Bechtel, W. & Mundale, J. (1999). Multiple realizability revisited. *Philosophy of Science*, 66, 175–207.

Bechtel, W. & Richardson, R.C. (1993). *Discovering Complexity*. Princeton, NJ: Princeton University Press.

Bedau, M. & Humphreys, P. (2008). *Emergence*. Cambridge, MA: MIT Press.

Bickle, J. (1998). *Psychoneural Reduction*. Cambridge, MA: MIT Press.

Bickle, J. (2003). *Philosophy and Neuroscience*. Dordrecht: Kluwer.

Bickle, J. (2006). Reducing mind to molecular pathways. *Synthese* 151, 411–434.

Brigandt, I. (*forthcoming*). Beyond reduction and pluralism. *Erkenntnis*.

Brigandt, I. & Love, A. (2008). Reductionism in biology. In E. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (*Fall 2008 Edition*), URL = <http://plato.stanford.edu/archives/fall2008/entries/reduction-biology/>.

Carnap, R. (1932*a*). Die physikalische Sprache als Universalsprache der Wissenschaft. *Erkenntnis* 2, 432–465. Transl.: *The Unity of Science*. London: Keagan Paul 1934.

Carnap, R. (1932b). Psychologie in physikalischer Sprache. Erkenntnis 3, 107–142. Transl.: Psychology in physical language. In A.J. Ayer (Ed.), Logical Positivism. New York, NY: Free Press, 165–198.

Carroll, J. (1994). Laws of Nature. Cambridge: Cambridge University Press.

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

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

Churchland, P.S. (1986). Neurophilosophy. Cambridge, MA: MIT Press.

Churchland, P.S. & Sejnowski, T. J. (1992). The Computational Brain. Cambridge, MA: MIT Press.

Craver, C.F. (2002). Interlevel experiments and multilevel mechanisms in the neuroscience of memory. Philosophy of Science 69, S83–S97.

Craver, C.F. (2007). Explaining the Brain. Oxford: Oxford University Press.

Feyerabend, P.K. (1962). Explanation, reduction, and empiricism. Minnesota Studies in the Philosophy of Science 3, 28–97.

Fodor, J. (1974). Special sciences: Or, the disunity of science as a working hypothesis. Synthese 28, 97–115.

Gillett, C. (2007). The metaphysics of mechanisms and the challenge of the new reductionism. In M. Schouten & H.L. de Jong (Eds.), The Matter of the Mind. Oxford: Blackwell, 76–100.

Hempel, C.G. (1949). The logical analysis of psychology. In H. Feigl & W. Sellars (Eds.), Readings in Philosophical Analysis. New York, NY: Appleton-Century-

Crofts, 373–384. Repr. in N. Block (Ed.), Readings in the Philosophy of Psychology, Vol. 1. Cambridge, MA: Harvard University Press 1980, 14–23.

Hendry, R. (2009). The Metaphysics of Chemistry. Oxford: Oxford University Press.

Hooker, C.A. (1981). Towards a general theory of reduction. Part I: Historical and scientific setting. Part II: Identity in reduction. Part III: Cross-categorial reduction. Dialogue 20, 38–59, 201–236, 496–529.

Hull, D. (1972). Reduction in genetics – biology or philosophy? Philosophy of Science 39, 491–499.

Jackson, F. & Pettit, P. (1992). Structural explanation in social theory. In D. Charles & K. Lennon (Eds.), Reduction, Explanation, and Realism. Oxford: Clarendon Press, 97–131.

Jaworski, W. (2002). Multiple-realizability, explanation and the disjunctive move. Philosophical Studies 108, 289–308.

Kim, J. (1992). Multiple realization and the metaphysics of reduction. Philosophy and Phenomenological Research 52, 1–26. Repr. in Kim, Supervenience and Mind. Cambridge: Cambridge University Press 1993, 309–335.

Kim. J. (1998). Mind in a Physical World. Cambridge, MA: MIT Press.

Kim, J. (2005). Physicalism, or Something Near Enough. Princeton, NJ: Princeton University Press.

Kripke, S. (1980). Naming and Necessity. Cambridge, MA: Harvard University Press.

Lewis, D. (1980). Mad pain and martian pain. In N. Block (Ed.), Readings in the Philosophy of Psychology, Vol. 1. Cambridge, MA: Harvard University Press, 216–222.

Lewis, D. (1994). Lewis, David: Reduction of mind. In S. Guttenplan (Ed.), A Companion to the Philosophy of Mind. Oxford: Blackwell, 412–431.

Machamer, P.K., Darden, L., & Craver, C. (2000). Thinking about mechanisms. Philosophy of Science 67, 1–25.

McCauley, R.N. (2007). Reduction: Models of cross-scientific relations and their implications for the psychology-neuroscience interface. In P. Thagard (Ed.), Handbook of the Philosophy of Psychology and Cognitive Science. Amsterdam: Elsevier, 105–158.

Molnar, G. (2003). Powers. Oxford: Oxford University Press.

Mumford, S. (2004). Laws in Nature. London: Routledge.

Nagel, E. (1961). The Structure of Science. London: Routledge.

Nickles, T. (1973). Two concepts of intertheoretic reduction. Journal of Philosophy 70, 181–201.

Oppenheim, P. & Putnam, H. (1958). Unity of science as a working hypothesis. Minnesota Studies in the Philosophy of Science 2, 3–36.

Pereboom, D. & Kornblith, H. (1991). The metaphysics of irreducibility. Philosophical Studies 63, 125–145.

Place, U. (1956). Is consciousness a brain process? British Journal of Psychology 47, 44–50.

Putnam, H. (1967). Psychological predicates. In W.H. Capitan & D.D. Merrill (Eds.), Art, Mind, and Religion. Pittsburg: Pittsburg University Press, 37–48.

Richardson, R.C. (1979). Functionalism and reductionism. Philosophy of Science 46, 533–558.

Richardson, R.C. (2007). Reduction without the structures. In M. Schouten & H.L. de Jong (Eds.), The Matter of the Mind. Oxford: Blackwell, 123–145.

Richardson, R.C. & Stephan, A. (2007). Mechanisms and mechanical explanation in systems biology. In F. Boogerd, F. Bruggeman, J. Hofmeyr & H. Westerhoff (Eds.), Systems Biology. Amsterdam: Elsevier, 123–144.

Rueger, A. (2006). Functional reduction and emergence in the physical sciences. Synthese 151, 335–346.

Sarkar, S. (1992). Models of reduction and categories of reductionism. Synthese 91, 167–194.

Schaffner, K. (1967). Approaches to reduction. Philosophy of Science 34, 137–147.

Schaffner, K. (1969). The Watson-Crick model and reductionism. British Journal for the Philosophy of Science 20, 325–348.

Schaffner, K. (1993). Discovery and Explanation in Biology and Medicine. Chicago, IL: University of Chicago Press.

Shapiro, L. (2004). The Mind Incarnate. Cambridge, MA: MIT Press.

Sklar, L. (1967). Types of inter-theoretic reduction. British Journal for the Philosophy of Science 18, 109–124.

Sklar, L. (1999). The reduction(?) of thermodynamics to statistical mechanics. Philosophical Studies 95, 187–202.

Stephan, A. (2006). The dual role of 'emergence' in the philosophy of mind and in cognitive science. Synthese 151, 485–498.

Walter, S. (2006). Multiple realizability and reduction: A defense of the disjunctive move. Metaphysica 9, 43–65.

Walter, S. (2008). The supervenience argument, overdetermination, and causal drainage: assessing Kim's master argument. Philosophical Psychology 21, 671–694.

Wimsatt, W.C. (1976). Reductionism, levels of organization, and the mind-body problem. In Globus, G.G, Maxwell, G., & Savodnik, I. (Eds.), Consciousness and the Brain. New York, NY: Plenum Press, 199–267.

Wimsatt, W.C. (2007). Re-Engineering Philosophy for Limited Beings. Cambridge, MA: Harvard University Press.

Woodward, J. (2003). Making Things Happen. Oxford: Oxford University Press.

Woodward, J. (2008). Mental causation and neural mechanisms. In Hohwy, J. & Kallestrup, J. (Eds.) Being Reduced. Oxford: Oxford University Press, 218–262.