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

Sven Walter and Markus Eronen

The goals of reduction have been numerous: revealing unity behind the appearance of plurality, showing that some phenomena are resultants of more fundamental phenomena, looking “downwards” into the composition of things in order to explain wholes in terms of their parts, or explicating one theory in terms of a more fundamental one.

Reduction is an idea as old as the human attempt to understand the world. 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; and current physicists are looking for the GUT, the “grand unified theory,” that will show that at very high energies the electromagnetic, the weak nuclear, and the strong nuclear forces are fused into a single unified field. Some of these projects are clearly ontological in nature (e.g., Leucippus and Democritus), others have a more methodological orientation (e.g., mechanism), and still others strive for theoretical simplification (e.g., Galileo and Newton, or the search for a GUT). Nevertheless, they may all be regarded as (attempted) reductions.

[*] We are grateful to Vera Hoffmann, Bob Richardson, and Dan Brooks for helpful comments on an earlier version.

Section 1 will provide a largely historical overview of the various philosophical accounts of reduction, focusing mostly on theory reduction, but taking into account ontological aspects of reduction as well. Section 2 considers how the issue of reduction has shaped recent philosophy of mind, in particular in connection with what is usually called “multiple realizability.” Section 3 then looks at some attempts to understand reductive endeavors in terms of “mechanistic explanations.” Finally, section 4 explores the interconnections between the idea of scientific reductions and the idea that our world is a layered one with distinguishable levels of organization.

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,” i.e., to find a common language into which all meaningful (i.e., verifiable) scientific statements could be translated. E.g., Rudolf Carnap (1932a, 1932b) and Carl Gustav Hempel (1949) argued that the language of physics may serve as a universal language of science. Scientific concepts, statements, and laws, if they are to have any meaning, they held, must be translatable into physical concepts, statements, and laws. Psychology, e.g., turns out to be “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; see also Ryle 1949). For example, Carnap (1932b, 170–171) argued, the psychological predicate “x is excited” is translatable into a physical predicate like “*x*’s body (especially his nervous system) 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.” His argument was that, first, the thesis of verificationism entails that predicates are synonymous iff they are applied on the basis of the same observations and, secondly, the second predicate simply enumerates the observations on the basis of which the first is applied. This translatability was also thought to lead to an *ontological reduction* of the property of being excited: since Carnap took properties to be the senses or intensions of predicates (see Carnap 1956), he held that a mental property M is identical to a physical property P iff the corresponding predicates “M” and “P” are synonymous.

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, in particular, psychology—were not translatable into physical language non-circularly: notoriously, someone who *wants* a beer will go to the fridge to get one, but only if she *believes* that there is beer in the fridge, does not *want* a whisky even more, does not *attempt* to stay sober etc. Cognitive synonymy (in the sense of identical confirmation conditions) or even predicate synonymy thus seemed to be too strong a requirement on both the theoretical reduction of psychology to physics and the ontological reduction of mental properties to physical properties. In the philosophy of science translational reduction was therefore replaced by more sophisticated models of reduction (considered below), while the philosophy of mind abandoned synonymy as a prerequisite for property-identities, paving the way for the modern idea that psychophysical property-identities are instances of 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 the special sciences like chemistry, biology, psychology etc. are reducible to more fundamental theories, eventually perhaps to a single “grand theory,” was retained. If higher-level theories were reducible to more fundamental theories, perhaps even to physics, then could not something like a “unity of science” be attained even if predicates like “*x* is soluble,” “*x* is a Chinese wisteria,” or “*x* is excited” are not translatable into purely physical terminology? Building on earlier work especially by Kemeny and Oppenheim (1956), 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 wholes which are fully decomposable into entities belonging to the universe of discourse of T_1. Oppenheim and Putnam’s approach faced some severe difficulties—e.g., 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 (see Sklar 1967)—but many of its key ideas are still visible in what came to be 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 seriously the idea that reduction consists in the *derivation* of the reduced theory T_2 from a reducing theory T_1. Such a derivation is possible, Nagel argued, if (1.) the terms of T_2 are connectable with the terms of T_1 (the "*condition of connectability*"), and (2.) given these connecting principles, all the laws of T_2 can be derived from the laws of T_1 (the "*condition of derivability*"). The connectability condition is essential because in interesting cases of reduction T_1 and T_2 are framed in partially disjoint vocabularies, so that without any connecting principles—or "*bridge-laws*"—the derivations required for reductions would be impossible. The exact nature of these bridge-laws, however, has been a matter of debate. Although Nagel allowed them to be material conditionals of the form "$\forall \underline{x}\ (F_{T1}\underline{x} \supset F_{T2}\underline{x})$" (see, e.g., Richardson 1979), it was usually assumed that biconditionals of the form "$\forall \underline{x}\ (F_{T1}\underline{x} \equiv F_{T2}\underline{x})$" are required to yield the ontological simplifications (viz., property-identities) that were seen as one of the main goals of reduction (see section 2).

In contrast to Oppenheim and Putnam's approach, Nagel's model of reduction was formally precise, but it also failed to fit standard examples of scientific reductions. For instance, Nagel was interested in the reduction of thermodynamics to statistical mechanics. Yet, he focused only on the derivation of the Boyle-Charles law ($pV = kT$) from statistical mechanics, pointing out that the derivation of thermodynamics as a whole would be immensely complicated—in fact, it is even more complicated than Nagel thought (see, e.g., Richardson 2007; Sklar 1999), and even the Boyle-Charles law is derivable from statistical mechanics only if a set of idealizing counterfactual assumptions is made. The possibility of a Nagelian reduction of thermodynamics to statistical mechanics is further diminished by the fact that central thermodynamical concepts like, e.g., “entropy,” are associated with a variety of distinct concepts in statistical mechanics, which do not exactly correspond to the concept of “entropy” in thermodynamics, neither separately nor taken together.

Another important problem for Nagel’s account proved to be the fact that the reducing theory often *corrects* the original theory, which means that the latter was *false*. E.g., 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. This is problematic because 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 reduction in science are possible or necessary. The majority of philosophers, however, responded by developing more sophisticated models (see, e.g., Causey 1977; Schaffner 1967), culminating in what has become known as, to use John Bickle’s (1998) term, “New Wave Reductionism.”

1.4 New Wave Reductionism

Like its precursor, *New Wave Reductionism* (NWR) is an allegedly universal model of reduction that sees reduction as a relation involving logical derivations between theories (see Bickle 1998, 2003; Hooker 1981, but also Churchland 1985; Churchland 1986;

Schaffner 1993). However, what is derived from T_1 is not T_2 itself, but an analogue or "equipotent isomorphic image" T_2a of T_2. This 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, then 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, then the amount of correction 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 (leading to a process resembling what Kuhn (1962) termed "scientific revolutions"). It is not clear, however, how exactly to evaluate the strength of the analogy between T_2 and T_2a (see Bickle (1998) for some suggestions, based on the structuralist/semantic view of theories). Additionally, NWR inherits two problems that already plagued early approaches to theory reduction.

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

Second, NWR retains the idea that the relata of reductions are formal or at least semiformal theories, phrased in either first-order predicate logic or in set-theoretic terms. Yet, some generally accepted cases of scientific reduction—for instance the reduction of genetics to molecular biology—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, including psychology or neuroscience, simply do not have any well-structured theories that could be handled logically or set-theoretically. Rather than trying to formulate such theories, these disciplines typically look for descriptions of mechanisms that can serve as explanations for patterns, effects, capacities, phenomena etc., and this explanatory enterprise at best involves fragments of formal theories (see Craver 2005, 2007; Cummins 2000; Machamer et al. 2000; McCauley 2007).

2. Multiple Realizability and Psychophysical Reduction

2.1 Multiple Realizability and Kim's Dilemma

In the philosophy of mind, the issue of reduction surfaces in the debate between reductive and non-reductive physicalists. Whereas the former hold that the mental can be reduced to the physical—at least ontologically, if not conceptually—the latter maintain that although ontological and conceptual reductions fail, mental properties are nevertheless not non-physical in any ontologically threatening (i.e., dualism implying) sense: the mental domain is irreducible, and thus ontologically and conceptually autonomous, but since the mental 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), Herbert Feigl (1958), Ullian Place (1956), and Jack Smart (1959) famously argued for a version of reductive physicalism according to which mental and physical properties can be identical even if mental predicates are not definable in purely physical terms: the thesis that consciousness is a brain process, Place (1956, 45) argued, is not a consequence of a successful conceptual reduction, but a "reasonable scientific hypothesis," on par with other theoretical identifications *a posteriori* like "Water is H_2O." The heyday of this early form of reductive physicalism was short-lived, however, for Putnam (1967) was quick to point out that the 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, in conspecifics, and even in one individual at different times. Fodor (1974) provided further support for anti-reductionism, arguing that Putnam's considerations apply, *mutatis mutandis*, 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 "$(\forall x)$ $(Mx \equiv Px)$" linking M with a physical property P will usually be false, and hence, since laws must be true, will fail to be *laws*. Statements like "$(\forall x)$ $(Mx \equiv (P_1x \vee \ldots \vee P_nx)$" linking S with the complete disjunction of all of its physical realizers will be true, but they, Fodor argued, cannot be laws either, because "$(P_1x \vee \ldots \vee P_nx)$" fails to designate a scientific kind (see section 2.2). Hence, there are no laws—and thus *a* *fortiori* no *bridge-laws* (at least no biconditional ones)—connecting special science properties with physical kinds, rendering Nagelian reductions of the special sciences impossible.

Due mostly to the arguments of Putnam and Fodor, non-reductive physicalism achieved an almost hegemonic status during the 1970s and 1980s. Jaegwon Kim (1992), however, forcefully argued that far from making psychophysical reductions impossible, multiple realizability actually engenders them. Those who try to support non-reductive physicalism by appeal to multiple realizability, Kim argued, face a dilemma, both horns of which lead to psychophysical reductions.

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 mental property M that is multiply realizable by physical properties $P_1x, \ldots, P_nx$ *via* the *disjunctive bridge-law* "$(\forall x)\ (Mx \equiv (P_1x \vee \ldots \vee P_nx))$." Call this the "*Disjunctive Move*." On the other hand, if Fodor is right and "$P_1x \vee \ldots \vee P_nx$" cannot designate a scientific kind, then, Kim argued, the mental predicate "M" cannot designate a scientific kind either. Therefore, if there are any mental laws at all, they must be couched in terms of the only law-fit predicates left, viz., "P_1," "P_2," …, "P_n," leading to so-called "local," "restricted," 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 mental property M iff x has physical property P. Call this "*Local Reductionism*." Either way, reductionism carries the day.

2.2 The Disjunctive Move

Even when the inadequacy of bridge-law based approaches to reduction was already evident in philosophy of science and Nagel's model had been replaced by more sophisticated intertheoretic models of reduction that did not appeal to bridge-laws, the debate between reductive and non-reductive physicalists was still centered on the availability of psychophysical bridge-laws, and it was not before the late 1990s that alternative models of reduction were finally explored (see section 2.4).

Consider, e.g., the *Disjunctive Move*. Its proponents assumed that the existence of bridge-laws linking mental and physical predicates is indeed sufficient for reductions, and then argued that multiple realizability is compatible with psychophysical reductions because there will always be true biconditionals linking a mental property M with the complete disjunction of its physical realizers. In response, opponents of the *Disjunctive Move* tried to show why such biconditionals cannot be bridge-laws, or at least why they usually fail to be bridge-laws.

Two important characteristics of laws are that they support *counterfactuals* and enable successful *predictions*, but biconditionals containing disjunctive predicates seem able to meet these conditions (see Kim 1992, 319; Owens 1989, 198; Seager 1991, 94). However, according to a traditional (though nowadays not universally accepted) view of laws, there are two other characteristic features of laws that have been taken to cause trouble. “$(\forall \underline{x})\ (F\underline{x} \equiv G\underline{x})$” is a law, it is sometimes claimed, only if it is (1.) *explanatory*, and (2.) *confirmed* by its positive instances (so that “F” and “G” are *projectible*, meaning 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 (see, e.g., Marras 1993;

Pereboom & Kornblith 1991; but see Jaworski 2002), and the predicate “$P_1x \vee \ldots \vee P_nx$” is unprojectible and does not designate a scientific kind because there is nothing significant in common from a causal point of view to all and only the individuals satisfying it: it is *causally heterogeneous* (see 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., *Local Reductionism*.

2.3 *Local Reductionism*

According to Kim, if the disjunction of a mental property M’s possible 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 nonnomic as a kind*? … It is difficult to see how one could have it both ways—that is, 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 psychophysical 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 from each other that the disjunction $P_h \vee P_r \vee P_m$ is causally heterogeneous, and thus unprojectible, and thus non-nomic. Given Kim’s argument that *having pain* cannot be nomic if $P_h \vee P_r \vee P_m$ is non-nomic (supposing, for the sake of simplicity, that the disjunction is complete so that “x has pain” and “$P_hx \vee P_rx \vee P_mx$” are coextensive, maybe even nomologically), there can thus be no laws about pain as such. The only projectible “pain-properties” suitable for laws 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). This results in restricted bridge-laws “$(\forall x)\ (S_hx \supset (Mx \equiv P_hx))$,” “$(\forall x)\ (S_rx \supset (Mx \equiv P_rx))$,” and “$(\forall x)\ (S_mx \supset (Mx \equiv P_mx))$” which sunder the psychological theory about pain in three different subfields, each of which is “locally reducible”:

> [I]f each of the psychological kinds posited in a psychological theory has a physical realization for a fixed species, the theory can be ‘locally reduced’ to the physical theory of that species, in the following sense. Let S be the species involved; for each law L_m of psychological theory T_m, $S \rightarrow L_m$ (the proposition that L_m holds for members of S) is the ‘S-restricted’ version of L_m. (Kim 1992, 328)

However, Kim eventually came to argue that *Local Reductionism* was based on a “seriously flawed notion of reduction” (Kim 1998, 12). A successful reduction of x to y, Kim (1998, 96) held, should be *explanatory* by making intelligible how x can arise out of y, and should *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, they are not explanatory:

even if “$(\forall \underline{x})$ ($\underline{x}$ has pain ≡ $\underline{x}$ has c-fiber firing)” were a law, this would not explain why having c-fiber firing should feel painish rather than, say, ticklish (Kim 1998, 95–96). Second, bridge-laws do not simplify ontology. One reason is that even if “$(\forall \underline{x})$ ($\underline{x}$ has pain ≡ $\underline{x}$ has c-fiber firing)” were a law, the properties *having pain* and *having c-fiber firing* could still not be identified because “$(\forall \underline{x})$ ($\underline{x}$ has pain ≡ $\underline{x}$ has c-fiber firing)” is contingent and its contingency can arguably not be blamed on a contingency involving an epistemic counterpart, as in all other cases of theoretical identifications *a posteriori* (see Kripke 1980). Another reason is that even if restricted bridge-laws like “$(\forall \underline{x})\ (S_h\underline{x} \supset (M\underline{x} \equiv P_h\underline{x}))$” are true and the predicate “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, viz., as 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).

Since Kim was convinced that only property-identities can yield the explanatory power and ontological simplification required for reductions, he tried to modify *Local Reductionism* in a way that preserved the key idea behind the dilemma’s second horn—that multiple realizability leads to species-specific reductions—while at the same time enhancing its species-relative coextensions of predicates into genuine, albeit species-relative, property-identities. The result was his model of *Functional Reduction*.

2.4 Functional Reduction

Kim's model of Functional Reduction (see also Jackson 1998; Levine 1991) is based on ideas from David Lewis (1980). Lewis had argued that instead of looking for psychophysical property-identities that hold across all possible worlds, we should identify mental properties with physical properties relative to species or structures. The concept "pain," Lewis claimed, is a functional concept of a state that occupies a certain causal role. Unlike typical functionalists, however, Lewis argued that this 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; emphasis added). 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 a functional predicate as usually conceived 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 properties (different physical occupants of the pain-role) in different species.

Given this, Lewis argued, if a mental predicate "M" means "the occupant of the M-role" and there is variation in what occupies the M-role, then not only the contingent laws relating "M" to physical predicates have to be restricted, the psychophysical property-identities themselves have to be restricted, too: "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, although restricted, property-identities, Lewis' account yields the ontologically simplifying and explanatory reductions Kim was asking 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 instead of some other physical property P*: having M-in-K just *is* having P.

Lewis-style reductions are essentially three-step procedures: A mental property M is first construed (*via* conceptual analysis) as the property characterized by a certain causal role; then the physical property P occupying the causal role definitive of M 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 is also the key idea of 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; see also 2005, 101)

This allegedly avoids the two problems that prevented bridge-laws from yielding genuine property-identities: (1.) Kripke's argument concerning the necessity of identities and (2.) the fact that a second-order property cannot be identified with a first-order property (see section 2.3). Regarding (2.), 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 functional property expressed by the gerund "being in pain"). Regarding (1.), Kripke's argument against the classical identity theory is ineffective against Lewis and Kim's approach, since it works only for identity statements containing rigid designators, whereas "*x* has pain" is supposed to be a *non-rigid* designator.

Kim's model of *Functional Reduction* is a kind of *eliminative reduction* (see section 1.4) in which the reduced property does not survive reduction. *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 (Kim 1999, 17). Although mental predicates and concepts may group physical properties in ways that are essential for descriptive, explanatory and communicative purposes, one will have to learn to live without universal mental properties like *having pain* (Kim 1998, 106). It is thus easy to see why Kim thinks the multiple realizability objection against reductionism fails: the differences among the physical realizers of a mental property M do not show that M is multiply realizable, but that the mental predicate “M” 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 within a single individual across time, so that *having pain* would be one physical property in Paul and another in Peter, or one physical property in Paul at t_1 and another at t_2. But further narrowing what appears to be a mental kind into ever more restricted physical structures seems theoretically self defeating, as with the increasing loss of generality the identifications will be theoretically uninteresting or even purely *ad hoc* (see Marras 2003, 257). According to the model of *Functional Reduction*, each difference in the physical implementation of a mental faculty, however small, entails that there are effectively different mental properties. Yet, brain-damaged patients seem to be able to compensate for losses in certain areas by *recovering* the mental faculties located in those areas elsewhere, suggesting that the *same* mental faculty can indeed be multiply realized by different physical structures.

The perhaps most objectionable consequence of the model of *Functional Reduction* is that it only allows for the reduction of properties definable in terms of causal roles. Properties which cannot be “primed” for reduction by construing them relationally or extrinsically—a notorious example are phenomenal properties like *having a reddish visual experience*, *having a lemonish gustatory experience* etc.—will not be susceptible to functional reductions. The problem is not only that some mental properties turn out to be irreducible, but that Kim's own *Supervenience Argument* (see, e.g. Kim 1998, 2005; see also Walter 2008) is designed to show that irreducible properties cannot be causally efficacious. Recently, Kim has thus reluctantly admitted that only mental properties that can be functionally reduced can be causally efficacious, while 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" (Kim 2005, 173).

Another important problem in the present context is that Kim presents the functional model as a realistic account of reduction in science (see, e.g., Kim 1998, 99), but has not shown that scientific cases of reduction actually conform to his model. Kim's main example is the reduction of the property of being a gene to strands of DNA, and even this example is presented only very schematically and in a way that does not justice to actual scientific practice, illustrating once again that discussions about reduction in the philosophy of mind have been largely unconstrained by, and 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 as the norm in science, at least not in fields like psychology, neuroscience, biology etc. 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" (see 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 the 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 are trying to formulate their explanations in terms of empirically discovered 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 William 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 (see, e.g., Bickle 2003; Craver 2002, 2007). A mechanistic explanation of memory consolidation describes the cellular and molecular mechanisms underlying it, i.e., it describes how the relevant parts of the memory system and their activities together result in the transformation of short-term into long-term memories. Central in 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 (e.g., when pre- and postsynaptic neurons are simultaneously active in certain parts of the hippocampus, the LTP mechanism results in the connections between them being enhanced, and this enhancement can last for days or weeks).

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, importantly, are not general levels of organization (see section 4), but important relative to this mechanism and perhaps different for other mechanisms: (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 is supposed to perform 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.).

Moreover, mechanistic explanations have both a "downward-looking" and an "upward-looking" aspect. In the LTP case, e.g., 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, compared to higher levels, it can be seen as supporting a kind of anti-reductionist "explanatory pluralism" (see Craver 2007; McCauley 2007; Richardson & Stephan 2007). This anti-reductionist conclusion receives further support from the "interventionist" account of causation (see Woodward 2003), according to which higher-level entities like, e.g., psychological states can have causal and explanatory relevance even though lower-level explanations in terms of implementing mechanisms are complete (see, e.g., Menzies 2008; Woodward 2008).

On the other hand, the process of "looking downward" and invoking parts of the mechanism to understand the behavior of the mechanism 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 (see, e.g., Bechtel 2008; Sarkar 1992; Wimsatt 1976). John Bickle (2003, 2006) has taken this reductive aspect of neuroscientific explanation seriously and argued for what he calls a "ruthlessly reductive" analysis of explanation in neuroscience. According to Bickle, when we look at the 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, there is nothing left, Bickle argues, for higher-level sciences like psychology to explain, and they are needed merely 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, however, 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 could possibly support Bickle's view. Therefore, it seems that mechanistic explanation pluralistically understood has a stronger case, and that explanation in neuroscience is, if at all, only ”somewhat” reductive—it is not “ruthlessly reductive,” and does not lead to eliminations of higher-level explanations.

4. Reduction and Levels of Reality

Talk of levels is ubiquitous, in science as well as in the philosophy of science and the philosophy of mind. 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 (see Craver 2007, 163–164).

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

A similar appeal to mereology can nowadays be found in Kim’s work with regard to levels of properties instead of levels of theories. The level of a property, Kim (1998, 92) argues, 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 (see 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 a criterion (see, e.g., 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 instance, is composed of smaller piles of snow, but this does not necessarily mean that the larger pile of snow is at a higher level than the smaller piles. 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 are black holes, or that raindrops are at the same level as are tiny computers.

Perhaps the most sophisticated, yet somewhat obscure, general and all-encompassing account of levels of organization has been developed by William Wimsatt (1994/2007). 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 that levels typically (but not necessarily) have. E.g., 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 (1994/2007, 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 instance, there are more regularities than at scales just slightly larger or smaller than the scale of atoms, 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 when we get to psychology and neuroscience, neat compositional relations break down. According to Wimsatt (1994/2007, 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 to one another like levels of organization. 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 1994/2007, 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.

Talk of levels also plays 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, such that the relata are mechanisms at higher levels and their components at lower levels (see Craver 2007, ch. 5). The notion of "level" at play here is in one respect fundamentally different from general levels of organization. Levels of mechanisms are not universal divisions in the structure of the world (á la Oppenheim & Putnam). Rather, different mechanisms have different hierarchies of levels. The levels in the spatial memory system, for instance, are different from those in the circulatory system. According to the mechanist, these local and case-specific levels are sufficient for understanding reductive explanation and interlevel relations in neuroscience (see, e.g., Bechtel 2008; Craver 2007). One limitation of this is, of course, 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.

References

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

Bechtel, W. & Richardson, R.C. (1993). Discovering Complexity. Princeton, NJ: Princeton University 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: Explicating the reductionism implicit in current cellular and molecular neuroscience. Synthese 151, 411–434.

Carnap, R. (1932a). 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.

Carnap, R. (1956). Meaning and Necessity, 2^{nd} ed. Chicago, IL: Chicago 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. (2005). Beyond reduction: mechanisms, multifield integration and the unity of neuroscience. Studies in History and Philosophy of Biological and Biomedical Sciences 36, 373–395.

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

Cummins, R. (2000). ‘How does it work?’ versus ‘What are the laws’. Two conceptions of psychological explanation. In F. Keil & R. Wilson (Eds.), Explanation and Cognition. Cambridge, MA: MIT Press, 117–144.

Feigl, H. (1958). The ‘mental’ and the ‘physical’. Minnesota Studies in the Philosophy of Science 2, 370–497.

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.

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.

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.

Jackson, F. (1998). *From Metaphysics to Ethics*. Oxford: Clarendon.

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

Kemeny, J.G. & Oppenheim, P. (1956). On reduction. *Philosophical Studies* 7, 6–19.

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. (1999). Making sense of emergence. *Philosophical Studies* 95, 3–44.

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

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

Kuhn, T. (1962). *The Structure of Scientific Revolutions*. Chicaco, IL: University of Chicago Press.

Levine, J. (1991). On leaving out what it’s like. In G. Humphreys & M. Davies (Eds.), *Consciousness*. Oxford: Blackwell, 121–136.

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.

Marras, A. (1993). Supervenience and reducibility: An odd couple. Philosophical Quarterly 43, 215–222.

Marras, A. (2003). Methodological and ontological aspects of the mental causation problem. In S. Walter & H.-D. Heckmann (Eds.), Physicalism and Mental Causation. Thoverton: Imprint Academic, 243–264.

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.

Menzies, P. (2008). The exclusion problem, the determination relation, and contrastive causation. In Hohwy, J. & Kallestrup, J. (Eds.) Being Reduced. Oxford: Oxford University Press, 196–217.

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.

Owens, D. (1989). Disjunctive laws? Analysis 49, 197–202.

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.

Ryle, G. (1949). The Concept of Mind. New York, NY: Barnes and Noble.

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

Seager, W.E. (1991). Disjunctive laws and supervenience. Analysis 51, 93–98.

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.

Smart, J. (1959). Sensations and brain processes. Philosophical Review 68, 141–156.

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. (1994/2007). The ontology of complex systems: Levels, perspectives and causal thickets. Canadian Journal of Philosophy (Suppl.) 20, 207–274. Revised reprint in Wimsatt, W.C., Re-Engineering Philosophy for Limited Beings. Cambridge, MA: Harvard University Press, 193–240.

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.