

Problems of Mental Causation - Whether and How It Can Exist

A Review of Jaegwon Kim's *Mind in a Physical World*

Rüdiger Vaas
University of Stuttgart
Posener Str. 85, 74321 Bietigheim-Bissingen
GERMANY

Ruediger.Vaas@t-online.de

Copyright (c) Rüdiger Vaas 2002

PSYCHE, 8(04), August 2002
<http://psyche.cs.monash.edu.au/v8/psyche-8-04-vaas.html>

KEYWORDS: mind-body problem, mental causation, free will, consciousness, functionalism, nonreductive physicalism, supervenience, multiple realization, reduction, emergence, explanation, properties.

REVIEW OF: Jaegwon Kim (1998). *Mind in a Physical World: An Essay on the Mind-Body Problem and Mental Causation*. Cambridge, Massachusetts: MIT Press. viii + 146 pp. ISBN: 0262112345, US\$ 25.00 (cloth), 0262611538, US\$ 15.00 (pbk, 2000).

ABSTRACT: There is a tension or even contradiction between mental causation - the belief that some mental events or properties are causally relevant for some physical events or properties - and the irreducibility of mental features to physical ones, the causal closure of the physical, and the assumption that there is no overdetermination of the physical. To reconcile these premises was a promise of nonreductive physicalism, but a closer inspection shows that it is, on the contrary, a source of the problem - namely, the unintelligibility of mental causation. This has to do with the widely-held assumption that the mental supervenes on the physical. How can the mental be causally relevant, then (because the physical seems to do all the causal work)? And what is the relationship of the mental and the physical (because supervenience must be explained)? There are many options, including identity, realizationism, emergence, or some kind of reducibility. But they all have their own problems, e.g. they threaten mental reality, the causal closure of the physical, or scientific explanations. All these aspects are covered in Jaegwon Kim's book *Mind in a Physical World* (1998). This paper is a detailed introduction to it, discussing and critically commenting it and those still intriguing, but also confusing and complicated issues of the mind-body problem, especially the ontology of mental causation.

"consciousness has plagued us and we can not shake it, though we think we're in control, though we think we're in control"
-Greg Graffin

1. Introduction

Analytic philosophy of mind has many parallels with chess: rigor and creativity in thinking, logical necessity and mental acrobatics, control of success and quality, transparency and an enormous complexity, effort as well as play, pleasure and a wish for more. A recommendable recent example for this is *Mind in a Physical World* by Jaegwon Kim. He is the William Perry Faunce Professor of Philosophy at Brown University, Providence, Rhode Island, and well known for his contributions about philosophy of mind in general and supervenience in particular (Kim 1993a, 1996). *Mind in a Physical World* is based on Kim's Townsend Lectures at the University of California, Berkeley, held in 1996 and extensively revised later. It is mainly about the mind-body problem, mental causation, reductionism, supervenience and emergence.

The book is rather short, clearly and elegantly written, but heavy-weighted, sophisticated, demanding and dense in its argumentation, thoughtful and sometimes provocative. It consists of a short preface, four chapters (120 pages altogether), notes (14 pages), references and a detailed index. Many points have already been stated earlier (in Kim 1992a and other essays reprinted in Kim 1993a for instance, see also Kim 1997a and 1997b) but are presented here within a larger context, including a more extensive discussion of other arguments and approaches. The book might be useful for advanced undergraduates in philosophy (particularly now that it is available in paperback), but is most appropriate and rewarding for graduate courses and readers who are familiar with the recent discussions in philosophy of mind and especially interested in ontology and conceptual foundations of the mind-body problem, mental and physical causation, functionalism, nonreductive materialism, supervenience, reductionism, emergence, levels of description and explanation, second-order properties, multiple realization, and the work of Lynne Rudder Baker, Ned Block, Tyler Burge, Donald Davidson, Jerry Fodor, Terence Horgan and John Searle, among others. Thus, it is probably too specialized and detailed for beginners or for more empirically oriented readers. Its importance and widely resonance is reflected already in an impressing number of reviews (cf., e.g., Crisp & Warfield 2001, Glymour 1999, Gorman 2000, Graham 2000, Hansen 2000, Heil 1999, Loewer 2001 and forthcoming, Lowe 2001, Marras 2000, Newman 2000, Polger 2000, Williams 2000).

The following four sections will introduce, review and briefly comment Kim's four chapters, while in the sixth section our perspective should be widened a bit.

2. The Relationship of Mind and Body

The first chapter, entitled "The Mind-Body Problem: Where We Now are", is a concise introduction into the main problems and their background. It starts with some short reminiscences of important roots of the current mind-body problem debates. For Kim, the mind-body problem is "finding a place for the mind in a world that is fundamentally physical" (page 2; all following page numbers are, if not stated otherwise, from Kim's book at issue). Its new roots are traced back to the late 1950s and early 1960s, especially to the work of Herbert Feigl, John Jamieson Carswell Smart and Ullin T. Place who had pioneered the (materialistic or physicalistic or naturalistic - these attributions are used interchangeable here) identity theory of mental and brain states. It was abandoned for the most part due to two principal objections - the multiple realization argument advanced by Hilary Putnam, and the anomalist argument by Donald Davidson -, leading to other (also mainly physicalistically conceived) concepts of mentality, namely functionalism and anomalous monism. Functionalism provided the new science of cognition with both a methodology and a metaphysics, while anomalous monism lead to a nonreductive form of physicalism and a strong autonomy of the (study of) meaning, intentionality, consciousness and normativity. There is some sort of a tension between these two approaches, and they are both based on assumptions which were central for the further debates in philosophy of mind and Kim's book at issue: Anomalous monism is founded on supervenience, and functionalism depends on multiple realizability.

2.1. Supervenience and Explanation

Supervenience is a difficult, complicated issue, because there are quite different definitions and forms (extensions) of it around (cf. Beckermann, Flohr & Kim 1992, Kim 1993a, Savellos & Yal  n 1995). But the main idea is easy to grasp: no mental difference without a physical difference; or (formulated via possible worlds): any two things that are exact physical duplicates necessarily are exact mental duplicates as well; or (taking the modal operator formulation): every mental property has a physical base that guarantees its instantiation, and without such a physical base a mental property cannot be instantiated. A more precise definition for (strong) mind-body supervenience is: "Mental properties supervene on physical properties, in that necessarily, for any mental property M, if anything has M at time t, there exists a physical base (or subvenient) property P such that it has P at t, and necessarily anything that has P at a time has M at that time" (9, cf. 39).

Usually, supervenience is understood as an asymmetric dependence (or determination): The mental is dependent on or determined by the physical. That is, however, not implied by the definition given above, because it "simply states a pattern of covariance between the two families of properties, and such covariances can occur in the absence of a metaphysical dependence or determination relation" (11). This criticism is *prima facie* justified but might be overcome with a better definition. (Besides, there is a tension between supervenience and identity, for supervenience is considered an asymmetrical

relation, while if M is identical to P it cannot asymmetrically depend on P or on itself - though Davidson, at least, has combined (weak) supervenience with (token) identity.) There is another problem which Kim also mentions *en passant* (123) without further elaborating it here: Supervenience is consistent with strange dualistic ontologies like epiphenomenalism, neutral monism (where the mental and the physical are two collateral effects of a single cause), a Spinozistic double-aspect theory, Malebranche's occasionalism or a Leibnizian parallelism for instance, where the mental is ontologically different from the physical, e.g. a separate substance, but cannot influence the physical. These possibilities would be precluded by the token identity of mental and physical properties or events together with supervenience. Token identity is a necessary component of Davidson's anomalous monism although Kim does not discuss it. (Note that token identity without supervenience does not imply materialism because it is also consistent with mentalism or idealism, i.e. the claim that everything is mental but some mental objects or events also have physical properties.) Supervenience alone is not sufficient for materialism. Thus, if token identity is too strong, as e.g. John Haugeland has claimed, there must be some other way to reject spooky metaphysics, e.g. accepting a principle of physical exhaustion. But Haugeland (1998, p. 119, cf. Vaas 2000a) argues that it is enough to get rid of "scientifically unmotivated, magically undetectable, and thoroughly bizarre" hypotheses by shifting the burdens of proof to the proponents of those hypotheses and accepting the heuristic rule "Don't get weird beyond necessity".

One of the advantages of supervenience is that it seems to give a clear meaning to the primacy of the physical domain and its laws without implying physical reductionism, thereby protecting the autonomy of the mental. This nonreductive physicalism/materialism still is probably "the most influential metaphysical position, not only on the mind-body problem but more generally on the relationship between higher-level properties and underlying lower-level properties in all areas" (8). One consequence of the entrenchment of this antireductionist consensus was the return of emergentism which flourished already in the 1920s (see 5.4. below).

However, supervenience does not give us a theory of mind-body relation. This is, as said above, because mind-body supervenience is consistent with a host of classic positions on the mind-body problem, including dualistic ones like emergentism, epiphenomenalism and even some sort of substance dualism. Anomalous monism is not sufficient either. Kim recognizes correctly that anomalous monism is a negative thesis: It tells us how the mental is not related to the physical -e.g. it is not (type-)identical with it nor explanatorically reducible to it nor are there lawful type-type connections -, but it says nothing about how the two are related. "Davidson's anomalous monism says no more about the relationship between the mental and the physical than the claim that all objects with a color have a shape says about the relationship between colors and shapes. [...] I believe we want our mind-body theories to tell us more, a positive story about how mental properties and physical properties are related, and hopefully also explain why they are so related" (5).

Kim concludes that "mind-body supervenience itself is not an explanatory theory; it merely states a pattern of property covariation between the mental and the physical and

points to the existence of a dependency relation between the two. Yet supervenience is silent on the nature of the dependence relation that might explain why the mental supervenes on the physical" (14). Thus, supervenience is only a kind of phenomenological but not metaphysically "deep" relation like causal dependence, reductive dependence, dependence grounded in defeasibility or entailment, and the like.

Nevertheless, supervenience plays an important role in the understanding of what is classically known as the *scala naturae* or layered world - the fact or at least our impression or pragmatic consensus that nature is stratified into different levels, orders, or tiers organized in a hierarchical structure. There are properties, activities and functions at each level that make their first appearance, or "emerge", at that level. This layered model can also be stated in terms of concepts rather than entities and their properties, i.e. different levels of descriptions, languages, analysis or explanation. The crucial question is: "How are the characteristic properties of a given level related to the properties at the adjacent levels - in particular, to those at the lower levels?" (16). - Now, when supervenience is superposed on the layered model, the idea of microindiscernibility results: "For any x and y , belonging to level L (other than the lowest level), if x and y are indiscernible in relation to properties at all levels lower than L (or, as we may say, x and y are microindiscernible), then x and y are indiscernible with respect to all properties at level L " (17). Supervenience here is mereological supervenience: the claim that properties of wholes are fixed by the properties and relations that characterize their parts, leading to a Democritean atomistic doctrine that the world is the way it is because the microworld is the way it is (18). Mind-body supervenience can be seen as an instance of mereological supervenience: a mental property M (a macroproperty) is supervenient on a certain physical mereological configuration P (certain microproperties, e.g. brain states).

But supervenience is not an explanation, thus important questions still remain: "Is M reducible to P in some appropriate sense? Can we explain why something has M in terms of its having P ? Are the P - M and other such supervenience relations further explainable [...] or must they be taken as brute and fundamental?" (18).

2.2. Reductionism and Multiple Realizability of Functional Second-Order Properties

Multiple (physical) realizability is a main topic in philosophy of mind as well as philosophy of science in general and one reason why neural correlates of consciousness are not enough for a full understanding of consciousness and mentality (cf. Vaas 1999). Multiple physical realizability means that mental properties are realized or implemented or executed by (or in) physical properties - analogous to abstract computational entities like Turing machines being realized e.g. in concrete electronic devices -, though they are neither identical nor reducible to them. Physicalist realizationism or functionalism states that no mental property can have a nonphysical realization - what is realized are second-order properties.

"F is a second-order property over set B of base (or first-order) properties iff F is the property of having some property P in B such that D(P), where D specifies a condition on members of B" (20). Thus, second-order properties are generated by quantification over the first-order properties (which are not first-order in any absolute sense but may well be second-order relative to another set of properties). Functional properties over B are "those second-order properties over B whose specification D involves the causal/nomic relation" (20). Whether a given property realizes a given functional property is an empirical question because the relation is contingent.

An example for this somewhat abstract terminology is dormitivity. A substance has this functional property just if it has a chemical property that causes people to sleep. Both Valium and Seconal have this second-order property, but it is based on different first-order (chemical) realizers - diazepam and secobarbital, respectively.

Multiple realizability of mental properties M means that M are taken as functional properties which could be realized by different first-order properties P, e.g. electrochemical properties of neural networks in natural brains, electromechanical properties of silicon-based devices in robots, not yet known properties of noncarbon-based intelligent extraterrestrials etc. Therefore, P has to be nomologically sufficient for M. "Thus, if $\langle P_1, \dots, P_n \rangle$ is a realization of $\langle M_1, \dots, M_n \rangle$, in the sense that each P_i is a realizer of M_i , it follows that the Ms are supervenient on the Ps. Physical realizationism therefore entails the supervenience thesis" (23). And physical realizationism, Kim's favorite position, explains the supervenience thesis: "the mental supervenes on the physical because mental properties are second-order functional properties with physical realizers" (24). Mental-physical correlations are also explained: By definition, to have M is to have a property with causal specification D, and in a certain system P is the property or one of the properties meeting specifications D.

This is a reductionist approach, because having M in these systems is nothing over and above having P. But these reductions are relative in two respects (25): First, in systems with different structure, the underlying mechanism realizing the reduced property may vary. Second, reductions remain valid only when the basic laws of nature are held constant - that is, only for nomologically possible worlds relative to the reference world. (Because the realization relation is relative to prevailing laws, the entailed supervenience thesis has no logical necessity, only nomological necessity; therefore, physically indiscernible systems in worlds with different laws do not necessarily instantiate the same mental properties.) Thus, accepting this approach, the reducibility of a property critically depends on its functionalizability and not on the availability of bridge laws between the different levels as it is assumed in Ernest Nagel's (1961) classical model of intertheoretic reduction. More on this below (5.2.).

In conclusion, while anomalous monism or supervenience fail to give a positive explanatory account of how mental properties are related to physical ones, physical realizationism does at least promise such an account. Its reductionist approach has the right form and content.

3. Problems of Mental Causation

The second chapter is about "The Many Problems of Mental Causation". The classical problem appeared with Rene Descartes' infamous interactionism: How could nonphysical minds influence physical matter? There was (and still is) no solution for this mystery and eventually Gottfried W. Leibniz and Nicolas de Malebranche chose to abandon mental causation at all to save substantial dualism. Descartes' problem is not our problem anymore - at least not the problem of naturalists. However, it turns out that the supervenience assumption and physical realizationism lead to some related problems which are similarly disturbing - Kim speaks of "Descartes's revenge against the physicalists" (46). The crucial question nowadays can be stated this way: "How is it possible for the mind to exercise its causal powers in a world that is fundamentally physical?" (30). Or, to put it like Terrence Horgan (1989): Is mental causation possible, i.e. a causal role of the mental qua mental?

For Kim, there are three doctrines currently on stage, each of which poses *prima facie* difficulties for mental causation: First, mental anomalism; second, computationalism and content externalism; and third, causal exclusion.

3.1. The Problem of Anomalous Mental Properties

Anomalous monism claims that there are no psychological laws. This seems to make mental causation impossible: "mental causation requires mental events to instantiate laws, but mental anomalism says there are not laws about mental events" (33). Mental properties have no causal role because they are completely dependent - or fixed - by physical properties. Davidson's position might even be compatible with a removal or strange redistribution of every mental property, leaving the causal structure of the world entirely untouched. Thus, mental properties are causally inert or epiphenomenal. Therefore, anomalous monism leads to the following problem of mental causation: "How can anomalous properties be causal properties?" (34).

There are three main routes to escape: First, change the underlying ontology, e.g. replace Davidsonian strict laws by laws which involve mental properties (a route taken e.g. by Jerry Fodor). Second, look to some sort of counterfactual dependencies rather than subsumptive causal laws to generate causal relations (a strategy put forward by Ernest LePore, Barry Loewer and Terence Horgan). Third, define a notion of causal relevance or efficacy weaker than causation regulated by strict laws (a more recent way gone by Davidson).

3.2. The Problem of Extrinsic Mental Properties

There are two sources of concern here. The first is computationalism or syntacticalism: It states that the shapes (syntax) of symbols, not their meanings (semantics), determine the course of computation. Thus, the representational content is not causally relevant. The second problem is content externalism: It states that semantic properties are relational or extrinsic, i.e. depend on the organism's history and ecological conditions. Thus, even if two organisms were physically identical regarding their intrinsic properties, due to the externalism of content they can differ in respect of the semantical properties they instantiate, i.e. the contents of their beliefs, desires, the truth conditions of their sentences etc. On the other hand, causative properties involved in behavior production are usually taken to be non-relational, or intrinsic, properties of the organism (of course they are caused, at least partly, from external influences). If inner states are implicated in behavior causation, it seems that all their causal work is done by their intrinsic (syntactic) properties, leaving their semantic properties causally otiose - semantical differences should make no difference to behavior output. Therefore, computationalism and content externalism lead to the following problem of mental causation: "How can extrinsic relational properties be causally efficacious in behavior production?" (37).

3.3. The Problem of Causal Exclusion

Finally, and this might be the main difficulty: What do mental properties M do, if physical properties P do all the causal work - especially if M is not reducible to P as stated by token physicalism? This problem of explanatory or causal exclusion arises if we assume that some form of physicalism is true, i.e. there are no violations of the causal closure of the physical domain (otherwise one would relapse into Cartesian interactionist dualism which puts physical and nonphysical events into a single causal chain). "If you pick any physical event and trace its causal ancestry or posterity, that will never take you outside the physical domain (Kim 1997a, p. 282). The principle of causal exclusion states that no event can have more than a single complete and independent cause, and the principle of explanatory exclusion states that no event can be given more than one complete and independent explanation (cf. 65) (For more precise notions of this principles and a defense against a recent criticism by Ausonio Marras 1998 & 2000 see the forthcoming paper by Andr  Fuhrmann and Wilson P. Mendon a). Therefore, causal exclusion of M leads to the following problem of mental causation: "Given that every physical event that has a cause has a physical cause, how is a mental cause also possible?" (38).

3.4. The Supervenience Argument

Next, Kim shows a dilemma that is implied even by the weakest form of physicalism, supervenience. The dilemma states: "If mind-body supervenience fails, mental causation is unintelligible; if it holds, mental causation is again unintelligible. Hence mental

causation is unintelligible" (46). This conclusion is the result of the following reasoning (39 ff):

- (1) Either mind-body supervenience holds or it fails.
- (2) If mind-body supervenience fails, there is no visible way of understanding the possibility of mental causation (if the physical closure principle is accepted).
- (3) Suppose that an instance of mental property M causes another mental property M* to be instantiated.
- (4) M* has a physical supervenience base P*. (Ex hypothesi.)
- (5) M* is instantiated on this occasion: first, because, ex hypothesi, M causes M* to be instantiated; second, because P* is instantiated on this occasion. (Note that the first "because" is meant in a causal, the second in a noncausal sense, for the instantiation of a subvening property does not cause the instantiation of the supervening one.)
- (6) M caused M* by causing P*. (Thus, mental-to-mental causation implies, or presupposes, mental-to-physical causation!)
- (7) M itself has a physical supervenience base P. (Ex hypothesi.)
- (8) P caused P*, and M supervenes on P and M* supervenes on P*.
- (9) The M-to-M* and M-to-P* causal relations are only apparent, arising out of a genuine causal process from P to P*.

Both M and P seem eligible as a sufficient cause for P* - thus an overabundance of causes, causal overdetermination, seems to occur. (Note that, pace John Searle, the relation between base properties and supervenient properties is usually not conceived as causal.) However, the causal power of M is based on the causal power of P. P causes P*, and M supervenes on P and M* supervenes on P*. That is mental properties seem to be parasitic on real causal, i.e. physical properties. Given that P is a full cause, there is no additional causal work left. So how could M cause anything if P already does all the work? How can M make any difference? - The situation is "like a series of shadows cast by a moving car: there is no causal connection between the shadow of the car at one instant and its shadow an instant later, each being an effect of the moving car. The moving car represents a genuine causal process, but the series of shadows it casts, however regular and lawlike it may be, does not constitute a causal process" (45).

This argument for the unintelligibility of mental causation, called the supervenience argument by Kim, is the result of superimposing mind-body supervenience on the causal

exclusion problem. It is one of the central problems discussed in the book, one of its highlights, and probably one of the main issues in recent philosophy of mind in general.

(Note that it is not entirely clear what the "unintelligibility" of mental causation should mean. Is it an epistemological claim, stating that "there is no visible way of understanding the possibility of mental causation" (40), cf. premise (2) above? Or are there also logical or ontological implications intended, i.e. an incompatibility or contradiction?)

3.5. Some Reflections on the Supervenience Argument

The supervenience argument has two parts. Part one: M-to-M causation implies or presupposes M-to-P causation, cf. (1)-(6). Part two: M-to-P causation is unintelligible, given mind-body supervenience, cf. (7)-(9). - Kim believes that there is a "real tension" (42) between the two instantiations in (5). They might be seen either in outright competition, or the truth of one sets conditions on the possibility of the other. But they put the claim of M to be a cause of M* in jeopardy: P* alone seems fully responsible for, and capable of accounting for, the occurrence of M*. For those who "do not see the tension", Kim offers a shorter route to the conclusion (42): There is a "plausible general principle" which suffices to justify the conclusion, namely: "To cause a supervenient property to be instantiated, you must cause its base property (or one of its base properties) to be instantiated". Part two of the supervenience argument leads either to causal overdetermination (i.e. if both M and P alone are sufficient for P*) or to a violation of the physical causal closure (if M is a necessary component in the causation of P*).

If the supervenience argument is cogent, it follows that at least one of its premises is false. Furthermore, the only option available to nonreductive physicalism will be that of rejecting mental-to-physical causation. This can be summarized in the following way (Hansen 2000, p. 470):

- (1) Suppose that a mental property instantiation M causes P*.
- (2) M has a physical supervenience base P.
- (3) On the standard accounts of causation P qualifies as a cause of P*.
- (4) Mental properties are not reducible to physical properties.
- (5) M and P are distinct (simultaneous) sufficient causes of P*.
- (6) Overdetermination is unintelligible.

Conclusion: Mental-to-physical causation is unintelligible given nonreductive physicalism. As it will be seen below, Kim's main attack goes against (4) and (5).

Another attempt to construe the supervenience argument is saying that a nonreductive physicalist cannot consistently espouse mental-to-physical causation if he believes in (1), (3) and (4) of the following premises (cf. Sturgeon 1998 and Hansen 2000, p. 473):

- (1) Nonreductive physicalism (property dualism),
- (2) Causal efficacy of the mental (mental-to-physical causation),
- (3) Principle of physical causal closure,
- (4) Unacceptability of (systematic) overdetermination.

The exclusion problem raises the worry that the conjunction of (1) and (2) brings with it the need to choose between rejecting (3) or (4), because (1)-(4) are jointly incompatible - any three will entail the negation of the fourth. (Note that this reconstruction does not make an appeal to mind-body supervenience explicitly, but supervenience is entailed in nonreductive physicalism, for otherwise nonreductive physicalism wouldn't count as physicalism.)

Thomas M. Crisp and Ted A. Warfield (2001) reconstruct the supervenience argument in a simplified manner and criticize it thoroughly. According to them, the argument goes as follows:

- (1) Either supervenience holds or it does not.
- (2) If it fails to hold, then, if property dualism and the causal closure of the physical ("closure" for short) are true, mental causation is unintelligible.
- (3) If it holds, then, if property dualism and closure is true, mental causation is unintelligible.

Conclusion: Mental causation is unintelligible if property dualism and closure is true.

First, Crisp and Warfield attack (2) by denying not only supervenience but also causal closure, which they define as "every caused physical event has a physical cause" (p. 305) in contrast to causal exclusion, which they define as "every caused physical event has only physical causes" (p. 307). Then it should be nomologically possible that M causes P* without being supervenient on P (it merely occurs together with P in some cases). This

would be impossible or inconsistent if causal exclusion holds. But according to Crisp and Warfield, this cannot be Kim's position here, because then supervenience would offer not even an initial glimmer of hope to the nonreductive physicalist who wishes to hold on to both mental to physical causation and physical causal closure; and whether or not supervenience holds, given property dualism, there cannot be mental to physical causation if causal exclusion holds (p. 308).

Second, Crisp and Warfield attack (3) by questioning causal or explanatory exclusion and the rejection of overdetermination. Kim's argument for (3) proceeds in two distinct steps: First, due to the principle of causal or explanatory exclusion "no event can be given more than one complete and independent explanation" (Kim 1993a, p. 239); thus a mental state M causes another mental state M* by causing its supervenience base P*, i.e. mental to mental causation implies mental to physical causation (let us ignore here whether M, strictly speaking, causes anything at all). Second, Kim argues that mental to physical causation is unintelligible given closure, supervenience and property dualism, because P appears to be a cause of P*, thus P* would be overdetermined and every case of mental causation would involve overdetermination; furthermore, if overdetermination is true and P would not occur, P* and M* would occur nevertheless due to M, and this violates the causal closure of the physical. - Regarding the first step, Crisp and Warfield (p. 310) question that both M and P* as explanations of M* violate causal or explanatory exclusion, because this principle requires that the competing explanations are independent, but M and P* are not. And even if the explanations are independent, they need not be conceived as competing, because they are fundamentally different kinds of explanations. This is true, but if one does not subscribe to certain views about causation (see 6.1), M seems to be causally superfluous for M*. Regarding the second step, Crisp and Warfield (p. 313) are not willing to give up overdetermination so quickly, but this seems to depend also on certain assumptions regarding causation.

Kim's supervenience argument is a surprising turn, because not long ago supervenience was the hope for saving the causal efficacy of the mental. Terence Horgan (1987), e.g., has argued for the physical supervenience of qualia to make them causally efficacious. And Jerry Fodor (1987, p. 18) was sure that "If mind/body supervenience goes, the intelligibility of mental causation goes with it." Now the contrary seems to be true (at least, to repeat it once more, if the causal closure principle is valid): Ironically, supervenience is not a solution but a source of the problem!

Are there ways out of the dilemma without introducing spooky free-floating mental forces interfering with the physical domain?

For the biological naturalist and mental realist John Searle (1995, p. 219, cf. Searle 2000, p. 173) the causal overdetermination is just the result of a confusion about different levels of description: "the same system admits of different causal descriptions at different levels all of which are consistent and none of which implies either overdetermination or failure of causal closure." - This might be a solution, but it depends on the ontology and language of causation which has to be made explicit. Besides, the idea of different causal descriptions of the same system raises the question whether all of these descriptions are

causally relevant (Meijers 2000, p. 181). Furthermore, the problem of mental causation is not, or not only, a problem of finding the right form or level of description. It is not just a matter of description whether or how mental properties or events cause physical ones. Of course one might say that there are psychoneural identities between M and P. Then M does per definitionem have physical effects (on P*) because it is, in a sense, a physical entity. But it is still the physical side of M-P which causes P* (or M*-P*). (There are similar problems of Donald Davidson's (1980, 1991, 1995) account who claims that causation is between events, not properties, making mental properties qua properties per definitionem epiphenomenal - but also physical ones.)

Another approach is to argue that mental causation is no more problematic than the causal properties of entities described in special sciences like chemistry or biology in relation to physics (see section 4.4. below). If we do not take them as causally inert, why should we give up mental causation?

4. Alternative Attempts

The third chapter is titled "Mental Causation: The Backlash and Free Lunches". Here, Kim discusses some of the strategies to answer the main question (or to escape from it): "if mental properties are physically irreducible and remain outside the physical domain, then, given that the physical domain is causally closed, how can they exercise causal powers, or enjoy any kind of causal relevance, in the physical domain?" (58f)

For Kim, "we cannot make the problem go away by making simple and inexpensive repairs here and there" (59). There are no cheap or even free lunch solutions, solutions at minimal philosophical costs, like to downplay the problem as a result of misplaced philosophical priorities or unmotivated metaphysical assumptions or a misunderstanding of the logic and metaphysics of causation or a false focus, description or starting point.

4.1. Explanatory Practice Versus Ontology?

Tyler Burge (1993) and Lynne Rudder Baker (1993), among others, argued that metaphysics has been given too much weight and explanatory practice too little (see also Meyering 2000 and Hardcastle). This might be true, but it does not dissolve the problem of mental causation which is a metaphysical problem. "It is the problem of showing how mental causation is possible, not whether it is possible, although of course what happens with the how-question may in the end induce us to reconsider our stance on the whether-question. [...] The issue is how to make our metaphysics consistent with mental causation, and the choice that we need to make is between various metaphysical alternatives, not between some recondite metaphysical principle on the one hand and some cherished epistemological practice or principle on the other" (61f). Furthermore, our practice of explanation is not independent of metaphysical assumptions. For example, if we accept

Davidson's argument that reasons are causes, we must necessarily deal with metaphysical issues (cf. section 6.1.). It might be true that intentional and physiological explanations need not and do not compete with each other if their premises are both true and consistent to each other. Nevertheless these explanations can be rivals if they both purport to causally explain a single explanandum. Then a tension results and we are compelled to ask how the two purported causes are related to each other.

One possibility is a compatibilism between the two kinds or levels of explanation. This is, of course, the approach of nonreductive physicalism where special sciences are somehow causally or theoretically autonomous in relation to physical theory. For instance, it has been argued that higher-order causal properties can cross-classify lower-order ones (Horgan 1997). However, if "mental properties and biological properties cross-classify basic physical properties, they cannot supervene on the latter" - and this "falls short of minimal physicalism" (69).

4.2. Counterfactuals

Another account is to base causal claims on counterfactual dependencies: given that *c* caused *e*, then we can say that if *e* had not occurred, *c* would not have occurred either. So does the truth of mental counterfactuals prove that mental causation exists, and make mental causes a difference which is not in conflict with physical explanations? - Suppose I go into a book store because I know that it sells a book about ancient Indians I want to buy. Now we can say that if I had not had that thought that I want to buy the book in that store, then, all other things being equal, I would not have entered the store; and, given that I did want to buy the book, then, all other things being equal, my entering was inevitable. The problem is that such an account will not make the need for further metaphysical clarifications go away. One reason is that the counterfactual account is consistent with epiphenomenalism (and other antinaturalistic ontologies): If my thought had not occurred, then I would not have been in some neural state *N*, and given that my thought did occur, *N* must have occurred, and this made it inevitable to enter the bookstore. Also, when *e*₁ and *e*₂ are collateral effects of a single common cause *c*, the counterfactual 'if *e*₁ had not occurred, *e*₂ would not have' and its converse can both be true, and given *e*₁, *e*₂ was inevitable. So, "what the counterfactual theorists need to do is to give an account of just what makes those mind-body counterfactuals we want for mental causation true, and show that on that account those counterfactuals we don't want, for example, epiphenomenalist counterfactuals, turn out to be false" (71).

4.3. Program Explanation

Another approach is to give up causal efficacy but save causal relevance of the mental. According to Frank Jackson and Philip Pettit (1990), 'G occurred because F occurred' can be a good, informative program explanation if (1) F is causally efficacious property with

respect to G, or if (2) F ensures (or programs for) the presence of some property P which is causally efficacious with respect to G, although F itself lacks causal efficacy in relation to G. (To "ensure" means "to program for" - like a computer program indicates certain events which go on at a lower, electronic level.) Take, e.g., the breaking of a fragile vase: Although fragility is not itself causally efficacious, and fails to be the cause for the breakage (because the cause is the molecular structure of the vase), it is nonetheless causally relevant in that it ensures the presence of a causally efficacious property (i.e. the molecular structure). - Depending on the notion of explanation, program explanation might save the explanatory power of mental descriptions, and special science properties in general, but it cannot be a causal explanation. We need not quarrel about the semantics, i.e. what an explanation really is - according to David Lewis (1986, p. 217) "to explain an event is to provide some information about its causal history" -, but program explanation gives up mental causation, and it is therefore an open question if a psychological explanation makes sense or is really needed anymore.

4.4. The Generalization Argument

Jerry Fodor (as well as Tyler Burge, Robert Van Gulick, Lynne Rudder Baker and others) has argued that if mental properties are epiphenomenal due to the supervenience argument, other properties have to be epiphenomenal too, namely special-science properties. This is because special sciences supervene on the basic science, e.g. chemistry supervenes on physics. However, "The causal laws of the special sciences and causal laws of basic sciences have in common that they both license ascriptions of causal responsibility" (Fodor 1989, p. 66). This is an important point, but it does not solve the problem: It is not clear whether special-science laws are causal laws on their own right or, if they are, whether mental causation is an exception nevertheless.

The problem, however, goes deeper. Fodor (1997, p. 160 f): "The very existence of special sciences testifies to reliable macrolevel regularities that are realized by mechanisms whose physical substance is quite typically heterogeneous. Damn near everything we know about the world suggests that unimaginably complicated to-ings and fro-ings of bits and pieces at the extreme microlevel manage somehow to converge on stable macrolevel properties. On the other hand, the 'somehow' really is entirely mysterious, and my guess is that that is what is bugging Kim." This "metaphysical mystery about functionalism" (Fodor 1997, p. 159) is how multiple realizability is possible, i.e. how can there be macroregularities that are realized by wildly heterogeneous lower level mechanisms. Robert W. Batterman (2000) recently offered a proposal based on the notion of universality in physics: Most of the details of the microstructure - those details that differentiate one realizer from another - are (contrary to Kim) irrelevant to the causal powers of the upper level property. Such similarities in the behaviors of diverse systems are well known from thermodynamics near critical points. Here, the microstructure, e.g. of a fluid, is largely irrelevant for describing the behavior of the particular system of interest; and many different systems with distinct microstructures exhibit identical behavior characterized by the same critical exponent.

Therefore, Batterman (p. 123) takes "multiple realizability as an instance of universality". Ned Block (1997, p. 120), too, has argued earlier that the realizers of the upper level properties, while heterogeneous, are not completely heterogeneous; laws of nature impose constraints on ways of making something that satisfies a certain description. But this does not solve the problem of mental causation nor does it prove or disprove the justification of the existence of special sciences. They are threatened by a generalization of the supervenience argument.

If the causal closure of the physical domain excludes the causal efficacy of mental properties in relation to physical properties, the same considerations might indicate that non-special-science properties could be causally efficacious with respect to their underlying lower-level properties. Thus, does the supervenience argument generalize? Is macrolevel causation in general a mere illusion? *Prima facie* this seems implausible, because then an errant baseball would not break a window and an earthquake would not cause buildings to collapse. Microphysics would be the only theory capable of generating causal explanations. But what if there is no causation on the microphysical level (standard interpretation of quantum theory claims the existence of acausal events, e.g. the radioactive decay of an atom)? And what if there is no bottom level at all?

Kim rejects the generalization argument, because it "does not have the full generality its supporters attribute to it. In particular, the exclusion-based worries about mental causation do not generalize across micro-macro levels" (84). It is important to distinguish between levels and orders, because the first-/second-/third-...order progression, i.e. the realization relation, does not track the micro-macro ordering but can apply to entities at a single micro-macro level. The reason for this is simply that both second-order properties and their first-order realizers are properties of the same entities and systems, i.e. are at the same level in the micro-macro hierarchy. E.g. a sleeping pill has both the second-order property dormitivity and the first-order chemical property which realizes dormitivity; the same goes for pain and its neural realizers. For something to have a second-order property is for it to have one or another of its realizers, that is, a first-order property satisfying the specification that defines the second-order property. Thus, the question whether something realizes a certain second-order property is independent from issues concerning micro-macro relations (82 f) - higher-level properties (which belong to a level above the micro level) are different from higher-order properties (which are specified by quantifying over a set of lower-order properties that satisfy a certain condition), e.g. functional properties. The order hierarchy does not track the levels hierarchy - both are, so to speak, orthogonal to each other - although many higher-level, first-order realizers are micro-based properties (but macro properties nevertheless). Therefore, "the supervenience argument does not have the effect of emptying macrolevels of causal powers and rendering familiar macro-objects and their properties causally impotent" (86). In conclusion, the generalization argument fails because it does not hold for micro-based properties like chemical or biological properties but only for mere supervening properties. (If it could be shown that the supervenience argument can be reformulated to threaten the efficacy of micro-based properties or that mental properties can be seen as micro-based properties, Kim's account is in trouble. Paul Noordhof (1999) tried to show this, but Kim (1999a) argued that he failed.)

This is an important result - and it shows a change in Kim's thinking who earlier (1993a, p. 96) has written that "all causal relations involving observable phenomena - all causal relations from daily experience - are cases of epiphenomenal causation", i.e. a mere reflection of some other underlying causal process. However, it does not save mental causation. The rejection of the generalization argument shows only that the causal exclusion problem is not an interlevel problem (which is solved by the order/level distinction). But there is still the problem of intralevel causal exclusion for which mental causation is an example. Thus, one needs another strategy to save mental causation.

5. Reduction

The fourth and final chapter has a promising title: "Reduction and Reductionism: A New Look". Terms like "reduction" or "reductionist explanation" are out nowadays, or even considered pejorative. Nevertheless, Kim hopes to persuade his readership "that reductionism about the mind is a serious, motivated philosophical position, and that although in the end we may decide to reject it, we should do so for the right reasons" (89). And Kim has much to offer for an unprejudiced investigation of reductions.

5.1. Troubles With Bridge Laws

First, Kim questions the account of Ernest Nagel who defined reduction as an inter-theoretic relation via bridge laws (conceived as biconditionals). They have the burden of linking the vocabulary of the theory targeted for reduction and that of the base theory, and thereby enable the derivation of the target theory from its reducer. One problem of this account is that unless we have two fixed and completed theories, we cannot say many useful things about what bridge laws may be needed for the reducibility of, say, mental states to physical states. (For a further discussion and many other concepts of reduction see, e.g., Vaas 1995, Van Gulick 2001.) But three important questions are already problems enough:

The availability question is raised by the multiple realization argument. If a higher-order property P has multiple realizers in lower-order properties, $Q_1, Q_2, \dots$, it is not possible to provide P with a single lower-order correlate Q to yield a biconditional bridge law ($P \leftrightarrow Q$). Thus P is irreducible to some single lower-order property. - There are two possible responses. First, the disjunctive strategy: If M is multiply realizable in, say, three ways, (P_1, P_2, P_3), one can take the disjunction ($P_1 \vee P_2 \vee P_3$) as M 's coextension in the base domain. Given that each of the P_i s is a realizer of M , it must hold with (at least) nomological necessity that $P_i \rightarrow M$ and $M \leftrightarrow (P_1, P_2, P_3)$. Second, species-restricted or local reductions might already be sufficient.

The explanatory question refers to the well-known explanatory gap or hard problem (cf. Chalmers 1996, Shear 1998): Why does a certain mental state occur (or a mental state at

all) when a certain neuronal activity occurs? Reduction must make it intelligible how certain phenomena arise out of more basic phenomena. This is a big challenge for physicalism. Kim states: "I believe physicalists should take the explanatory question seriously. It isn't that on physicalism every phenomenon must be physically explainable. [...] We may not be smart enough, diligent enough, or live long enough. But if a whole system of phenomena that are *prima facie* not among basic physical phenomena resists physical explanation, and especially if we don't even know where or how to begin, it would be time to reexamine one's physicalist commitments" (96).

The ontological question is whether a certain reduction actually simplifies matters. The price might be too high if the addition of the bridge laws as new basic laws of the base theory and new descriptive terms expand both the language and ontology of that base theory. Nagel's account is not helpful here, as it is even compatible with emergentism and many other forms of dualism like double-aspect theory, epiphenomenalism, or parallelism, where mental phenomena might be linked to physical phenomena via bridge laws. Kim therefore concludes that "the question whether or not mentality is Nagel-reducible via bridge laws to the physical cannot be a significant metaphysical issue" (97).

5.2. A Functional Model of Reduction

A real ontological simplification by means of reductions would enhance bridge laws, $M \leftrightarrow P$, into identities, $M = P$. "Identity takes away the logical space in which explanatory questions can be formulated" (98).

If both M and P are intrinsic properties and the bridge law connecting them is contingent, an identification is impossible because if $M = P$ is necessary, $M \leftrightarrow P$ cannot be contingent (assuming Saul Kripke's (1980) widely accepted thesis that M and P are rigid designators). However, taking M as a relational or extrinsic property, a functional model of reduction can be applied by constructing M as a second-order property - or more precise and less misleading: as a second-order description of properties, or a second-order concept - defined by its causal role. Such a functional construal serves as an explanation of why the correlation $M \leftrightarrow P$ holds and as a ground for a metaphysically contingent identity $M = P$. This identity is only nomologically necessary (it holds in all nomologically possible worlds in relation to the reference world, i.e. in all worlds with the same laws of nature); it is not necessary tout court. So here comes the central question (101): "Is the mental amenable to the kind of functionalization required for reductive explanation, or does it in principle resist such functionalization?"

If the functionalist conception of the mental is correct for all mental properties - i.e. not only intentional phenomena but also phenomenal qualities of experiences (qualia) -, then mind-body reduction is in principle possible. - This might be a surprising result, because functionalism counts as the principal contemporary form of mind-body antireductionism. Kim's account is exactly opposite to this. For him, the functionalist conception of mental properties is required for mind-body reduction. "In fact it is necessary and sufficient for

reducibility. If this is right, mind-body reductionism and the functionalist approach to mentality stand or fall together; they share the same metaphysical fate" (101).

This is an important conclusion. Although functionalism appears to conflict with the identity-theoretic reduction of the mental to the physical - at least with type identity physicalism - because of multiple realization, Kim proposes that functional properties can be identified with physical attributes of their realizers: Every instance of M is identical with some instance of the domain of base properties, e.g. P1, P2 etc. This means that functional properties are nothing but physical properties. In effect, Kim urges that we should embrace physical monism in order to solve the problem of mental causation.

Functional properties, as second-order properties, do not bring any new causal powers into the world (beyond the causal powers of their first-order realizers). However, for Kim this does not mean that they are causally impotent - contrary to Ned Block (1990) who takes them as epiphenomenal. Interestingly, it is Block to which Kim attributes the idea that mental properties are "second-order". If they are multiple realizable, they are causally heterogeneous. Thus, according to Kim's view, mental properties are not causally inert if they are functionally reducible; "the functionalization of mental properties enables them to escape the supervenience argument" (116). Difficulties arise for those mental properties that resist functionalization, e.g. qualia, or more general for supervenient properties that are not reducible to their base properties. If they do exist and supervene on physical properties, the problem of mental causation is not solved for them.

By the way: If M has multiple realizers, say P1 and P2, it should not be understood as a disjunctive property ($M = P1 \vee P2$) but simply as "Having M = having P1 and P2", i.e. not as a single disjunctive explanation but a disjunction of two explanations. The functional reduction of M requires the functionalization of M and consists of identifying each instance of M with its realizer P_i relative to the species or structure under consideration (and relative to the reference world). M is P1 in species 1, P2 in species 2 etc. Given that each instance of M has exactly the causal powers of its realizer on that occasion, all the causal/explanatory work done by an instance of M, occurring by virtue of the instantiation of realizer P_i , is done by P_i (110).

This is an elegant move, but it has its price. George Graham (2000, p. 549) for example worries "that the realizations are not just physically diverse but perhaps open (unpredictable, indeterminately diverse in physical type), that mental/functional properties may be embodied in physically unprincipled or endless arrays of organisms, systems or devices. [...] The determinate oneness of the mental cannot be identified with the indeterminate manyness of the physical." This is, however, no threat for the identity theory if supervenience holds. But Kim admits that mental properties cease to be alike in their own right. They are "sundered into their diverse realizers in different species and structures, and in different possible worlds" (111). Thus, M can no longer be a single, unified property. Mental states like pain are no natural kinds; there are different kinds of pain - the pain of a human, a lizard or a Martian, for instance. (By the way, David Lewis (1980) has defended a similar species-specific functionalistic type identity theory long ago.) Sceptics like Graham (2000, p. 550) interpret this as evidence for "irrealism about

the mental": "As I understand mental-property ascription, mental predicates can isolate patterns of causal activity across physically dissimilar domains. [...] So mental properties cannot be sundered into their realizers, as Kim claims." However, irrealism of transspecific mental kinds does not imply the non-existence of mental properties as such, and a heterogeneity of instances - an ontological issue - does not preclude abstract descriptions - an epistemological or pragmatic issue. What lends unity to the talk about functional properties is, according to Kim (110), "conceptual unity, not the unity of some underlying property".

5.3. Causal Inheritance and the Short-Coming of Nonreductive Physicalism

One of the main points in *Mind in a Physical World* is "that under nonreductive physicalism it is not possible to make sense of mental causation [...] under realization physicalism, it turns out that not only mental properties, but also causal relations in which mental properties figure, must be physically realized. If this is right, mental properties can have no causal powers beyond the causal powers of their underlying physical realizers. This result is at odds with the claims of nonreductive physicalism that mental properties are distinct from physical properties (if "distinct" means anything here, it must be "causally distinct"), and that the special sciences are in the business of constructing causal explanations at "higher" levels which are not "visible" bottom up" (Kim 2001).

Nonreductive realization physicalism runs into difficulties with mental causation on account of a principle that Kim (1992a) finds unavoidable: The principle of causal inheritance. It states: "If a second-order property F is realized on a given occasion by a first-order property H (that is, if F is instantiated on a given occasion in virtue of the fact that one of its realizers, H, is instantiated on that occasion), then the causal powers of this particular instance of F are identical with (or are a subset of) the causal powers of H (or of this instance of H)" (54). Thus, F can have no causal powers going beyond those of its realizer H. This is simply an implication of the definition of (functional) second-order properties (see 5.2.).

Earlier, Kim (1993a, p. 327) reasoned that instances of M that are realized by the same physical base must be grouped under one kind, since ex hypothesi the physical base is a causal kind. And instances of M with different realization bases have to be grouped under distinct kinds, since - also ex hypothesi - these realization bases are distinct as causal kinds. Given the multiple realization of mental kinds, i.e. that they can be realized by diverse physical causal kinds, it follows that mental kinds are not causal kinds. And hence they are disqualified as proper scientific kinds. (For Kim also assumes a causal individuation principle: Kinds in science are individuated on the basis of causal powers, i.e. objects and events fall under a kind, or share a property, if they have similar causal powers.) Nonreductive physicalists like Terrence Horgan (1996, p. 602 f) are not convinced about this, because they assume that mental causal kinds can "cross-classify" the physical causal kinds that realize mental kinds in a given species of creature:

Although two instances of a mental property *M* have dissimilar causal powers *qua* physical (they are distinct physical causal kinds, for their realization bases are distinct as physical causal kinds), they nonetheless have similar causal powers *qua* mental (thus, the two token states fall under a common mental causal kind: *M*). Suppose for example that *M* appears in robust psychological laws about humans, and that there occur two instances of *M* in humans with different physical realization bases. Then, although these two token states have dissimilar causal powers at the physical level of description, they nonetheless have similar causal powers at the mental level of description (in so far they both are *M*-instances and thus both fall under common *M*-involving psychological causal laws about humans). Therefore, Horgan argues that Kim's conclusion cited above is false even if all the premises are true.

5.4. Emergence

Whether the mental is amenable to the kind of functionalization required for reductive explanation or whether it resists such functionalization is an hotly debated issue. A main contrahent to reductionism is emergentism. Like "reduction" the term "emergence" is complex and ambiguous (cf., e.g., Vaas 1995, Stephan 1999, Van Gulick 2001). Most emergentists accept that emergent properties are determined by basal conditions, but they deny that the basal conditions can explain why it is just these emergent properties emerge from them. Thus, they deny the functionalizability of the properties they claim to be emergent. For the emergentists these properties are intrinsic, i.e. properties in their own right. However, explanation - also an ambiguous term - and epistemic restrictions are not the whole issue. A stronger form of emergence also denies the causal inheritance principle. Here, it is assumed that the emergent properties have their own distinctive causal powers that are irreducible to those of their basal conditions - some have even postulated the existence of "downward causation" (cf. Campbell 1974, Meyering 2000, Popper & Eccles 1977, p. 19, Rockwell; see also Kim 1992b for a different meaning). But then it is difficult to see if or how emergence is in agreement with the causal closure of the physical world.

The existence of emergent properties refutes physicalism only if emergence comes along with a breakdown of supervenience. If emergence is a failure of the explainability of the supervenience relation due to a second component missing, then physicalism stays alive, because explainability is not an ontological issue. However, explainability is difficult to specify. Are we supposed to appeal to additional laws of nature in order to explain the co-variation and dependence between two properties? Or do we have to give a metaphysical grounding of the supervenience relation (e.g., *A*'s supervene on *B*'s because they are *B*'s, made up of *B*'s, caused by *B*'s or have a common cause with *B*'s)? And is an epistemic notion like explainability needed at all to characterize an essentially metaphysical position like physicalism? (Cf. Kim 1993a, p. 76: "[T]he thesis that a given domain supervenes on another is a metaphysical thesis about an objectively existing dependency relation between two domains; it says nothing about whether or how details of the dependency relation will become known so as to enable us to formulate explanations,

reductions, or definitions.") The paradigm of self-organization is a promising approach to explain the emergence of novel properties and complex processes based on nonlinear dynamics, phase transitions, chaos theory, synergetics etc. without a breakdown of supervenience. And this can even shed some light on the explainability of supervenience. Alexander Rueger (2000, p. 479), e.g., has proposed to supplement the supervenience relation by imposing a requirement of robustness which is motivated by the notion of structural stability familiar from dynamical systems theory: "A supervenience relation between property classes in a system is grounded (or explainable) if (i) the relation is structurally stable or robust; or (ii) if the relation is unstable, the instability occurs in a stable way, i.e., the system belongs to a family of systems which, as a family, is structurally stable." The second case might count as a case of diachronic property emergence. But how this approach might enhance our understanding of the mind matter relation remains an open issue.

According to Kim (1992b, p. 126), emergentists hold that the relationships between emergent properties and their basis are essentially inexplicable (this is, of course, only one of many notions of emergence). Reductions require over and above supervenience an explanation of why a Nagelian bridge law holds. This lead Andreas Hüttemann and Orestis Terzidis (2000, p. 276) to the following definition of emergence: A property M of a system x is emergent with respect to the property P of the same system if (1) M supervenes on P and (2) the (restricted) bridge law $M \leftrightarrow P$ is inexplicable. But this is not a satisfying approach because in most cases which are classically called emergence there are no bridge laws; furthermore, bridge laws are, according to Nagel (1961), essential for reductionism (but between theories, not properties). And, as Hüttemann and Terzidis (2000, p. 279) recognized, "since [better: as far as] it is impossible to give the required explanation of bridge laws" emergence as defined above "seems to be a relation that holds necessarily as long as a supervenient relation holds between two sets of disjunct properties of one and the same system" what makes it impossible "to divide the class of physical systems into interesting subclasses". Thus, Kim's functional model of reduction is much more promising, and it might also be a better indicator for emergence, namely in cases where it fails. (Section 6.1. has more on antireductionism and emergence, cf. also Kim 1999b.)

5.5. Qualia

Kim notes that he has "nothing essentially new to offer" about qualia and admits doubts that the functionalist account can sufficiently handle them (102); "if emergentism is correct about anything, it is more likely to be correct about qualia than about anything else" (103). More recently Kim (2001) said that "the discomfort of qualia epiphenomenalism can be substantially allayed if not completely eliminated. I think that it is the absolute, intrinsic aspect of a quale that is not functionalizable (and hence, according to me, irreducible) and epiphenomenal; that is, it is the greenness of a green quale or the redness of a red quale - or the fact that green looks like this or red looks like that - that is not functionally identifiable. However, that a green quale and a red quale

look different can be functionalized." For example, the ability to pick out ripe tomatoes from a mound of lettuce leaves does not depend on intersubjectively sharing the same qualia space, only on the ability to reliably distinguish red from green. "Thus, qualia differences and similarities are functionalizable and hence causally efficacious, although the intrinsic "looks" and "feels" of qualia may not be."

It should be added that mental content just might be irreducibly subjective because of our sensory structure - we are systems with centered information acquisition - and specific "mineness" qualia as an effect of proprioception, creating a first-person perspective. But this is an epistemological, not an ontological issue. Furthermore, as Michael Tye (1995) has shown, there is a fundamental ambiguity in our notion of "facts": Some facts depend on our familiarity with them, i.e. on sensory experience and the use of phenomenal concepts based upon this experience; and some facts are intersubjective and therefore independent from this kind of phenomenal experience. According to Tye, the difference may be the result of two different modes of presentation. It is an epistemic and conceptual difference, but not an ontological one. (Cf. also Pete Mandik's (2001) more recent account of the subjectivity of consciousness by explicating the ways in which mental representations may be perspectival. For another new approach to show how qualia might be causally effective and functionally relevant, see Llinás 2001, cf. Vaas 2001a).

Last but not least it should be noted that, according to Ron Chrisley (reviewed by Vaas 2002a, p. 75), scientific objectivity is neither a view from nowhere, nor are scientific explanations logical deduction. Scientific objectivity as "the view from nowhere", i.e. missing any perspectivity, underlies Thomas Nagel's (1974, 1980) account of the incompatibility of subjective and objective: Consciousness is a subjective phenomenon, only accessible from a subjective point of view, while scientific accounts should be objective, contain no subjectivity and therefore cannot explain it. However, science does not need a perspectiveless perspective but is rather a way of negotiating human, perspective-bound views; it is "a view from anywhere". Therefore, objectivity and subjective experience are not incompatible. Scientific explanations as logical deduction seems to underlie David Chalmers' (1996) argument for a lack of logical (!) supervenience. He presupposes that scientific explanations must show how the lower-level facts logically entail what is to be explained. For example, we can imagine a Zombie Earth, physically identical to earth, but with no qualia at all. Thus, consciousness is not logically entailed by physical facts and cannot be explained by them. However, science is not purely deductive (and it is a subject of change). Ron Chrisley argued that showing how low-level facts entail what is to be explained is only one mode of explanation. An explanation need only make it intelligible how something with one description also has another description. Scientific understanding consists in a practical capacity to interrelate the two descriptions.

5.6. Bad News and the Remaining Options

So where are we now? The remarkable last section of Kim's book, entitled "The Options: Good News and Bad News", is pretty explicit and radical. He begins: "If we are prepared to go for a functionalization of all mental properties, we will be embracing an all-encompassing reductionism about the mental, and this will solve the problem of mental causation. That's the good news. On a reductionist position of this sort, however, the causal powers of mental properties turn out to be just those of their physical realizers, and there are no new causal powers brought into the world by mental properties. Many will consider that bad news. But the real bad news is that some mental properties, notably phenomenal properties of conscious experiences, seem to resist functionalization, and this means that there is no way to account for their causal efficacy within a physicalist scheme. These properties are not able to overcome the supervenience argument" (118f).

But there are even more bad news to come. Ultimately, "all roads branching out of physicalism may in the end seem to converge at the same point, the irreality of the mental" (119). The alternative to physicalism, perhaps even worse, is that we are forced to stumble into the "uncharted territory" of dualism - with "little knowledge of what possibilities and dangers lurk in this dark cavern" (120). If Kim is right, the currently popular middle-of-the road-positions like property dualism, anomalous monism, and nonreductive materialism, are not easily tolerated by robust physicalism. Physicalism has its price.

So what are the options?

One either stays with physicalism and accepts at least the principle of causal closure. Or one abandons physicalism in favor of a serious form of dualism (not just epiphenomenalism) and rejects at least mind-body supervenience. Another option, not discussed by Kim, is to subscribe to a form of idealism or panpsychism. If one stays with physicalism, there comes another choice: Either one retains mental properties that are supervenient and yet irreducible, i.e. not functionalizable, e.g. qualia, but accept their causal impotence. Or one embraces mental eliminativism and denies the reality of these irreducible properties altogether. And - more bad news - there may really not be much difference between these two options, because the possession of causal power is a plausible criterion for distinguishing what is real from what is not. But something that "has nothing to do, no purpose to serve might as well, and undoubtedly would in time, be abolished" as already Samuel Alexander (1920, p. 8) said decades ago. In this respect, eliminativism and epiphenomenalism both come close to the same: mental irrealism. Thus, if one wants to stay with physicalism and save mental properties and their causal efficacy, the reductionist alternative, i.e. functionalizing both intentional as well as phenomenal mental properties, seems to be the only option. But is this option much better? "Doesn't it lead to the conclusion that the mental has no distinctive role of its own, having been entirely absorbed into the physical domain? That again may seem to some as a form of mental irrealism, and one might think it makes no sense to save mental causation while relinquishing mentality as a distinctive reality" (119).

6. One Step Back For A Broader Perspective

Before going into some details, it might be useful to characterize and compare some of the opponents on stage, that is some of the main ontological theories of mind-matter-relationship. To get a - somewhat rough - overview it is helpful to sketch a classification or taxonomy relating to the following (non-exhausting!) features: substance dualism (sd), causal closure of the physical (cp), independent downward causation (dc), monistic property ontology (po), causal compatibility (cc) or coexistence of mental causation via nonphysical properties with physical causation even if the physical is causally closed, reducibility of psychology (rp), global supervenience (gs), and multiple realizability (mr). If a position has such a feature, there is a plus sign (+) in Table 1, otherwise a minus sign (-); some positions are either compatible with both possibilities or come along in different variants.

Table 1

Comparison between some ontological theories about the relationship between mind and matter. The abbreviations are explained in the paragraph of the main text above.

Ontology	sd	cp	dc	po	cc	rp	gs	mr
Eliminative Materialism	-	+	-	+	-	-	(-)	-
Identity Theory	-	+	-	+	-	+	+	+/-
Monistic Reductive Materialism	-	+	-	+	-	+	+	+
Dualistic Reductive Materialism	-	+	-	-	+	+	+	+
Nonreductive Physicalism	-	+	+/-	-	+	-	+	+
Functionalism	-(+)	+	-	+/-	+/-	+/-	+	+
Anomalous Monism	-	+	-	-	+	-	+	+
Epiphenomenalism	+/-	+	-	-	-	-	+	-
Traditional Emergentism	-	-	+	-	-	-	+/-	-
Cartesian Interactionism	+	-	+	-	-	-	+/-	-
Leibnizean Parallelism	+	+	-	-	+	-	+	-

Nonreductive physicalism promised a - nowadays quite popular - middle ground position which holds that mental properties are real, higher- level and causally efficacious without violating physical laws or the causal closure of nature. This turned mental causation into mainly an epistemological issue, i.e. a matter of explanations, not ontology. And the supervenience relation was welcomed as tight enough to capture the essence of physicalism but loose enough to be compatible with the irreducibility of the mental to the physical. In *Mind in a Physical World* however, Kim argues that nonreductive

physicalism cannot keep its promises, thus one either has to swallow reductionism or to abandon physicalism, turning the problem back into an ontological one.

6.1. Mental Causation and Ontology

The problem of mental causation arise from the incompatibility of four apparently plausible propositions (cf. Loewer 2001, p. 315):

(1) Mental causation: Many or all mental events cause physical events; or mental properties are causally relevant for some physical properties.

(2) Irreducibility: Mental events resp. properties are distinct from physical ones.

(3) Causal closure: The physical is nomologically and causally complete or closed, i.e. there are no nonphysical causes. (Indeterministic acausal fluctuations, e.g. quantum processes, do not violate the principle of causal closure.)

(4) No overdetermination: Physical events are not pervasively causally and/or nomologically overdetermined.

Taken together, these four propositions together are inconsistent, thus at least one of them must be dropped. Nonreductive physicalism is committed to (1), (2) and (3), which implies the denial of (4). But according to Kim, (4) is more plausible than the conjunction of the others, so one of them has to go. The reason for this is that, due to Kim, on the one hand overdetermination is implausible and may come into conflict with (3) which, on the other hand, makes mental causes dispensable "because they are parasitic on real causal processes" (45).

The basic theses of nonreductive physicalism, applied to the mental, are (according to Kim 1993a, p. 344, cf. Clarke 1999, pp. 296 f):

(1) Physical monism: All concrete particulars are physical.

(2) Antireductionism: Mental properties are not reducible to physical properties; no mental property is nomologically or metaphysically coextensive with a physical property (property dualism).

(3) Physical Realizationism: All mental properties are physically realized; that is, whenever an organism or system instantiates a mental property M, it has some physical property P such that P realizes M in organisms of its kind; thus, mental properties supervene on their realizing properties.

(4) Mental Realism: Mental properties are real properties of objects and events; they are not merely useful aids in making predictions or fictitious manners of speech.

(5) Alexander's Dictum: To be real is to have causal powers.

According to Kim (see also Clarke 1999, p. 300), nonreductionism must hold the different causal powers thesis: Each irreducible mental property carries with it causal powers that are different from those of any base property on which that mental property depends. Therefore, nonreductionism must reject the causal inheritance principle (5.3.) - if the instance of M is understood as a property instance and not an event. But nonreductionism holds a causal realization principle (Kim 1993a, p. 352): If a given instance of S occurs by being realized by Q, then any cause of this instance of S must be a cause of this instance of Q (and of course any cause of this instance of Q is a cause of this instance of S). Nonreductionism, then, is committed to mental-to-physical causal transaction (sometimes called downward causation), if it is supposed that mental events cause anything at all. This means to "accept emergent causal powers: causal powers that magically emerge at a higher level and of which there is no accounting in terms of lower-level properties and their causal powers and nomic connections" (Kim 1993a, p. 326). However, although emergentism holds that irreducible mental properties carry with it causal powers that no one of its realizing, i.e. base properties carries, it is not committed to the view that mental properties carry no underived causal powers - each mental property might derive its causal powers simply from all or most of its base properties. Thus, emergence is not committed and in fact denies weird causal powers which act against physical laws. It might even be compatible with a modification of the causal inheritance principle, namely what Randolph Clarke (1999, p. 310) has called event-power inheritance principle: "If M is instantiated on a given occasion by being realized by P, then what this exemplifying of M can cause, in the circumstance, is just what (or perhaps a subset of what) this exemplifying of P can cause." Of course this sketch has to be worked out in much more detail and it is obviously a matter of the background ontology. Furthermore, emergence still runs into troubles because of the principles of causal and explanatory exclusion. How can overdetermination be avoided? "If the exemplifying of P suffices for the exemplifying of P*, then, even if any causation by the exemplifying of M depends on this underlying physical causation, mental causation still seems redundant, unnecessary, to be causing the same thing twice over. [...] if no exemplifying of a mental "property" can do anything new, then no mental "property" is real [according to Alexander's Dictum]" (Clarke 1999, pp. 313 ff). - It is also noteworthy that Kim frames the metaphysical conception of nonreductive physicalism solely in terms of mind-body supervenience. But, as he emphasizes (14), supervenience is not a metaphysically deep relation and must be supplemented with a metaphysical relation that grounds or accounts for it. Therefore, "it seems that an assessment of the prospects of non-reductive physicalism needs to be made in light of what such relations are available to the non-reductive physicalist, and not mind-body supervenience alone" (Hansen 2000, p. 470).

It should be evident already that the whole discussion about mental causation, irreducibility, emergence, causal closure etc. involves a number of ontological assumptions about events, properties, causation, laws of nature etc. This is beyond the scope of this review, but it must be stressed nevertheless that as long as there is no clarification and no consensus among the opponents regarding formal ontology, chances are low to get the mind-body problem and mental causation straight.

Many philosophers would say that what causes, and what is caused, are events. Yet there is considerable disagreement on how finely events are individuated. (For a discussion see Elder (2001), who argues in contrast to Kim that mental causation faces no competition from the microphysical level). For example, Davidson's anomalous monism rejects (2) for events while accepting the rest, and rejects (1) for properties and maintaining the others. For him, causation applies only to events and it makes no sense to require a causal role for properties while critics point out that his properties are epiphenomenal. For Kim, an event is an instantiation of a property by an individual or a relation between some individuals at a time, and properties are not just descriptions (pace Davidson) but features of nature. According to Davidson (1991, 1995), causation is ultimately a relation based on fundamental physical laws (which we do not have yet, however, and possibly never will!) while mental events or properties are not in any strict lawful relation to physical ones and therefore anomalous. Davidson (1991, pp. 12 f) writes: "It is events that have the power to change things, not our various ways of describing them. Since the fact that an event is a mental event, i.e. that it can be described in a psychological vocabulary, can make no difference to the causes and effects of that event, it makes no sense to suppose that describing it in the psychological vocabulary might deprive the event of its potency. [...] For me, it is events that have causes and effects. Given this extensionalist view of causal relations, it makes no literal sense [...] to speak of an event causing something as mental, or by virtue of its mental properties, or as described in one way or another." Kim (1993b, pp. 21 f) insists that properties (resp., more precisely, their instances) are causally efficacious: "The issue has always been the causal efficacy of properties of events - no matter how they, the events or the properties, are described [...] we also need a way of talking about the causal role of properties [...] [T]he causal relation obtains between a pair of events because they are events of certain kinds, or have certain properties". For Kim, causation is a relation in which the cause produces the effect in an unequivocal manner, therefore his causal exclusion principle ("Don't multiply causes beyond necessity"). Barry Loewer rejects this account of causation and tries to strengthen and improve the counterfactual account which Kim (see section 4.2. above) has rejected. Loewer (2001, p. 324): "In this case a free lunch is preferable to indigestible metaphysics." And Ausonio Marras (1998, 2000) asks how cause and effect are individuated: as a nonextensional, explanatory relation ("in-virtue-causation") or as an ontological one? This is crucial to answer the following question: Did *c* cause *e* in virtue of *c*'s being *M* and *e*'s being *M** (resp. *P**), or did *c* cause *e* in virtue of *c*'s being *P* and *e*'s being *M** (resp. *P**) (where *M* and *M** supervene on *P* and *P**, respectively)?

Peter Menzies (forthcoming) thinks that the view of the causal inefficacy of mental states relies on false assumptions and a subtle misunderstanding of the concept of causation. He argues that we conceptualize causation not as a categorical, absolute relation, but as

entities occupying certain functional roles, defined with respect to abstract models. Therefore, there can be different levels of causation which need not be in competition with each other. "There may be a level at which mental states cause behaviour by way of distinctive psychological pathways; and a different level at which physical brain states cause behaviour by way of distinctive neural pathways." Menzies denies Kim's assumption of causal exclusion, i.e. that with the exception of cases of overdetermination, no event has more than one complete causal history. Multiple causal pathways seem coincidental in cases of overdetermination but unnecessarily redundant in cases of mental causation, because in contrast to merely accidental cases of overdetermination (e.g. a car accident caused both by an icy road and faulty brakes even if one of those would have been sufficient), mental causation would be strictly and law-like due to the supervenience relation. But according to Menzies, this claim of an unnecessary duplication "unsatisfactorily begs the question in presupposing that one causal pathway suffices for the explanation of a phenomenon, making other causal pathways explanatorily redundant. This line of thought fails to recognise that our claims about the causation of behaviour are made relative to models and that different models involve different kinds of abstraction that shape the identification of different, non-competing causal processes." Menzies proposed a concept of causation which allows that a single event (e.g. raising an arm) "can have two different complete causal histories". He argues that this is not simply a kind of overdetermination, because of the supervenience relation underlying mental causation and because "multiple causes in the examples of mental causation are picked out within different models, whereas the multiple causes of overdetermination examples are picked out within the same model." It is beyond the scope of this review to summarize and discuss Menzies' concept of causation. However, an opponent might argue that what is at issue here is not the coexistence of different models or levels of description, but the ontology of causation. If, say, causation is not (only) a logical relationship but some physical influence - e.g. transformation of energy, momentum and other quantities covered by a law of conservation -, and if some sort of physicalism is true, then there is only one sort of causes, say microphysical ones, and higher-level or downward causation is just a convenient abstraction, a description of pragmatic and heuristical value but quite far away from what "really" happens. Furthermore, there is the question, whether "causal powers involve laws", as Kim (1993a, p. 327) believes; if so, it has to be shown that mental causation is covered by such laws (or that a model of it entails such laws) - something against which, e.g., Donald Davidson (1980) has argued vehemently. Thus, Menzies might hit the causal exclusion assumption not on the level where Kim applies it. But of course it can be replied that causation as physical influence is just another model or concept and perhaps not even a contradicting one. Thus, to repeat it once more, it is necessary to get our ontological concepts straight. This might be an endless task, for it is always possible to change the rules of the game and play differently again and again.

Other questions are related to the nature of properties. Kim's functional model of reductions "rejects the commonly-held view that functional accounts of mental properties are not reductive accounts of those properties. The related claim that functional properties may not be at a higher level than the properties on which they supervene is also contrary to the common view" (Newman 2000, p. 89). And how can second-order properties be identical to first order-properties? Kim answers this with a form of a semantic ascent by

replacing the talk of second-order properties with the talk of second-order predicates or concepts or descriptions (104). But now there are different-order predicates which designate nevertheless the same property - one expressing a role-concept, the other a role-filler concept. How can they be coreferential? And how can a property be both realized by and identical with a given property? Like supervenience/dependence but unlike identity, realization is an asymmetrical relation: The mental is realized by the physical but not vice versa. Thus, how can realization be the same relation as identity? Kim's answer is that there is no unitary, functional property M over and above each of the realizers; M is not the disjunctive property $P1 \vee P2$ but it means simply to have $P1$ or $P2$ (see 5.2.); there are no disjunctive properties. For Ausonio Marras (2000) this is "incongruous with the very spirit of functionalism. [...] The search for unification and nomological homogeneity is what was supposed to drive functionalism; if we go for 'local' reductions and 'sunder' the multiply realized properties into their diverse realizers, much of this homogeneity is lost." This is true albeit there is at least some conceptual unity as Kim has emphasized (110). Perhaps Kim's approach is more incongruous with the letter than with the spirit of functionalism, because a more abstract kind of unitary psychology despite physiological diversity seems possible nonetheless, and advocates of functionalism took species-specific constraints into consideration long ago (cf. Lewis 1980). But there might be other approaches, e.g. models of functional analyses, whose aim is not to identify functional properties with physical ones but to explain how they are realized in physical systems (cf. Cummins 1983, ch. 2), e.g. by analyzing the configurations of lower-level components whose properties and mode of organization enable the system to play the functional roles of the functional properties (cf. Marras 2000). This might also avoid the problem of type-identifying mental properties with their species/structure-specific realizers.

In conclusion, ontological assumptions play a crucial role, and the situation is, at the moment, confusing at least or even a mess. It seems that a convincing solution of the problem of mental causation requires progress and some consensus in formal ontology, including the notion of different levels of reality. "Kim puts the ball squarely in the court of those who favor a multi-tiered view of reality. By insisting that all parties lay their ontological cards on the table. Kim has made it more difficult for philosophers of mind - and indeed anyone attracted to a layered picture of the world - to keep ontology at arm's length" (Heil 1999, p. 772 f). Furthermore, if Kim's supervenience argument (3.4.) is correct, mental causation is unintelligible. But then even Kim's favored physical realizationism does not help because it entails the supervenience thesis (23). Unintelligibility of mental causation is not sufficient for denying its existence - if we take mental causation as an ontological, not an epistemological (explanatory) or pragmatic issue -, but might suggest that direction.

6.2. The End of the World?

If the real causes are physical causes, what room is there for the mental to act? Clark Glymour (1999, p. 459) has distinguished four main strategies to show that it is reasonable to believe in mental causation nevertheless:

(1) Deny the Premise Strategy: The claim that the mental supervenes on the physical is false.

(2) Humpty Dumpty Strategy: The argument against mental causation is sound but for a long time we have been talking very satisfactorily about thoughts as causes, and we plan to continue, and science has nothing to do with it.

(3) Functionalist Strategy: Mental states are causal dispositions, implemented or realized by physical states.

(4) Identity Strategy: Some mental entities are identical to physical entities, the very same thing, like the Morning Star and the Evening Star. If those physical entities can be causes, then so can those mental entities.

Strategy (1) would kill the causal closure of the physical world (a strategy taken by Cartesian dualism and sometimes in the context of quantum philosophy and the search for quantum correlates of consciousness, but see Vaas 2001c). A modified combination of (3) and (4) is Kim's approach, but it is not clear if it does work. If other versions of (3) and (4) won't work either, (2) remains. Of course, the problem of mental causation dissolves if we give up mental causation. Would this loss really be so terrible? For Jerry Fodor (1989/1990, p. 156) at least it would: "If it isn't literally true that my wanting is causally responsible for my reaching, and my itching is causally responsible for my scratching, and my believing is causally responsible for my saying [...], if none of that is literally true, then practically everything I believe about anything is false and it's the end of the world."

This sounds fairly exaggerated. But if mental causation is only an illusion this could indeed mean the end not only of Libertarian free will (which is not discussed by Kim and seems to be an unintelligible or even incoherent wish anyway, cf. Vaas 2001b & 2002b) but also of humans as agents and cognizers. This is because "the possibility of human agency evidently requires that our mental states - our beliefs, desires, and intentions - have causal effects in the physical world: in voluntary actions our beliefs and desires, or intentions and decisions, must somehow cause our limbs to move in appropriate ways, thereby causing the objects around us to be rearranged"; and "the possibility of human knowledge presupposes the reality of mental causation: perception, our sole window on the world, requires the causation of perceptual experiences and beliefs by physical objects and events around us. Reasoning, by which we acquire new knowledge and belief from the existing fund of what we already know or believe, involves the causation of new belief by old belief [...] If you take away perception, memory, and reasoning, you pretty much take away all of human knowledge" (31). Thus, the problem of mental causation

threatens human agency and knowledge while the problem of determinism and scepticism, respectively, threatens only one of them.

6.3. Social Affairs and the Intentional Stance

At the risk of destroying the world we should nevertheless question these implications. For if it would inevitably turn out that there is indeed no mental causation in any relevant sense, wouldn't the life of most of us stay just the same? (Some wouldn't believe the conclusion, others - e.g. because of religious convictions - believe in determinism anyway.)

Furthermore, there is an evolutionary argument why it is rewarding to believe in mental causation, agency, free will and the like (Vaas 2000b, 2001b, 2002b): The intentional stance (Dennett 1988), i.e. ascribing intentional states to others, necessarily includes ascribing volitions to them and assuming that they have the power to transfer their volitions into actions somehow, because this is the only way to get advantages from the intentional stance at all. For, if other beings are thought to have intentions which are causally inert, this ascription of intentions and hence volitions simply wouldn't matter. However individuals endowed with the intentional stance are better prepared for the struggle of social life. And it is advantageous to assume the volitions of others as somehow being (essentially) independent of the environment or the past, because this makes it a lot easier to deal with them due to the fact that complex organisms can act (or react) quite differently in similar circumstances and quite similar in very different circumstances. Thus, the intentional stance is not an irrelevant luxury but a powerful tool to get along with the complexity of the social world. And this was, as it seems (Byrne and Whiten 1988, Whiten and Byrne 1997), a significant selective pressure for the rapid evolution of the higher primate's bigger brains and large intelligence, including their elaborated mental abilities like representation of complex social relationships, higher-order intentional stance, mind-reading, and primitive theory of mind - which allows sophisticated degrees of co-operation, deception and defense against deceptions. Since better access to food or a safer place to sleep or a higher rank in the complex hierarchies of primate societies normally increase the probability of producing more offspring than other group members, social intelligence pays off pretty well (Vaas, 2002c).

There is further reason to take a concept of volition as evolutionarily advantageous, and this is just the other side of the coin: To deal with other individuals in a complex way also means to plan one's own actions carefully in an explicit way and evaluate their effects. This presupposes some kind of awareness of one's own volition, hence a concept of will and self. Higher-order representations also take one's own mental states into account - not only for decisions and follow-up analyses but also as a parameter in the plans of others regarding oneself. Thus, it is reasonable or even necessary to ascribe volitions to oneself, too - because otherwise one cannot reason about the mental states of others who are presumably dealing with oneself. This makes one's own volitions explicit - and much more flexible. The concepts of mental causation, volition, actions and self-notions have

been flourishing at least since the point from which there has been language with an inbuilt grammatical structure distinguishing between subjects and objects, active and passive, present and future - but probably much earlier.

6.4. Good News At Last

Of course, *Mind in a Physical World* will not be the final word in philosophy of mind. It is no turning point either, but an important provisional result - and, like science, all philosophy which does not dead-lock itself, is provisional. Neither is this book Kim's last word. In the near future - and this sounds like good news - he will publish his Daewoo Lectures held in Seoul, five lectures titled Taking Physicalism to the Limit. Kim (2001) says: "I am trying to set down, I hope for the last time, my views on the mind-body problem and mental causation [...] I believe I have reached a more or less stable view on the issues, a view I feel comfortable with. My general message is: Physicalism, strictly speaking, is false, but it is the truth near enough, and near enough is good enough!"

Finally, there is some more good news which *Mind in a Physical World* carries, although it does not even mention it: While books such as this are neither meant nor able to solve all the problems at issue, they shape and enhance them, putting forward future research and clearing as well as improving the way to proceed. This is more than one can usually expect from a philosophical book. In any case Kim has shown how strenuous as well as exciting analytic philosophy of mind still is - and here we may return to the analogy with chess from the beginning. The mind-body problem and mental causation are not easily solved. There's still work to do for all contrahents. And, like in chess, there are many good ideas and moves needed - for all parties in the game.

Acknowledgements

For many helpful comments and suggestions I am very grateful to André Spiegel.

References

- Alexander, S. (1920). *Space, Time, and Deity*, Vol. 2. London: Macmillan.
- Baker, L. R. (1993). Metaphysics and Mental Causation. In Heil, J. & Mele, A. (eds.) (1993), *Mental Causation*. Oxford University Press. Oxford, pp. 75-95.
- Batterman, R. W. (2000). Multiple Realizability and Universality. *The British Journal for the Philosophy of Science*, 51, 115-145.

Beckermann, A., Flohr, H. & Kim, J. (eds.) (1992). *Emergence or Reduction?* Berlin, New York: de Gruyter.

Block, N. (1990). Can the Mind Change the World? In Boolos, G. (ed.), *Meaning and Method*. Cambridge: Cambridge University Press, pp. 137-170.

Block, N. (1997). Anti-Reductionism Slaps Back. *Philosophical Perspectives*, 11, 107-132. Online at <http://www.nyu.edu/gsas/dept/philo/faculty/block/papers/AntiReductionism.html>

Burge, T. (1993). Mind-Body Causation and Explanatory Practice. In Heil, J. & Mele, A. (Eds.) (1993), *Mental Causation*. Oxford: Oxford University Press, pp. 97-120.

Byrne, R. & Whiten, A. (1988). *Machiavellian Intelligence*. Oxford: Clarendon Press.

Campbell, D. T. (1974). 'Downward Causation' in Hierarchically Organized Biological Systems. In Ayala, F. J. & Dobzhansky, T. (Eds.), *Studies in the Philosophy of Biology*. London, Basingstoke: Macmillan, pp. 179-186.

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

Clarke, R. (1999). Nonreductive Physicalism and the Causal Powers of the Mental. *Erkenntnis*, 51, 295-322.

Crisp, T. M. & Warfield, T. A. (2001). Kim's Master Argument. *Noûs*, 35, 304-316.

Cummins, R. (1983). *The Nature of Psychological Explanation*. Cambridge: MIT Press.

Davidson, D. (1980). *Essays on Actions and Events*. Oxford: Oxford University Press.

Davidson, D. (1991). Thinking Causes. In Heil, J. & Mele, A. (eds.) (1993), *Mental Causation*. Oxford: Oxford University Press, pp. 3-17.

Davidson, D. (1995). Laws and cause. *Dialectica*, 49, 263-279.

Dennett, D. (1988). The intentional stance in theory and practice. In Byrne, R. & Whiten, A. (1988), *Machiavellian Intelligence*. Oxford: Clarendon Press, pp. 180-202.

Elder, C. L. (2001). Mental Causation versus Physical Causation: No Contest. *Philosophy and Phenomenological Research*, 62, 111-127.

Fodor, J. (1987). *Psychosemantics*. Cambridge: MIT Press.

Fodor, J. (1989). Making Mind Matter More. *Philosophical Topics*, 17, 59-79. Reprinted in Fodor, J. (1990). *A Theory of Content and other Essays*. Cambridge: MIT Press, pp.137-159.

Fodor, J. (1997). Special sciences: Still autonomous after all these years. *Philosophical Perspectives*, 11, 149-163.

Fuhrmann, A. & Mendonça, W. P. (forthcoming). Explanatory exclusion and causal relevance. Online at <http://www.uni-konstanz.de/FuF/Philo/Philosophie/Fuhrmann/papers/xxclusion.pdf>

Glymour, C. (1999). A Mind is a Terrible Thing to Waste - Critical Notice: Jaegwon Kim, Mind in a Physical World. *Philosophy of Science*, 66, 455-471.

Gorman, M. (2000). Jaegwon Kim's Mind in a Physical World. *The Review of Metaphysics*, 54, 937-938.

Graham, G. (2000). Book Reviews. *The Philosophical Quarterly*, 50, 548-551.

Hansen, C. M. (2000). Between a Rock and a Hard Place: Mental Causation and the Mind-Body Problem. *Inquiry*, 43, 451-491.

Hardcastle, V. G. On the Matter of Minds and Mental Causation. Online at <http://www.phil.vt.edu/Valerie/6014/6014papers/MMMC.html>

Haugeland, J. (1998). *Having Thought. Essays in the Metaphysics of Mind*. Cambridge, London: Harvard University Press.

Heil, J. (1999). Review: Jaegwon Kim. Mind in a Physical World: An Essay on the Mind-Body Problem and Mental Causation. *The British Journal for the Philosophy of Science*, 50, 769-773.

Horgan, T. (1987). Supervenient Qualia. *Philosophical Review*, 96, 491-520.

Horgan, T. (1989). Mental Causation. *Philosophical Perspectives*, 3, 47-76.

Horgan, T. (1996). Kim on the Mind-Body Problem. *British Journal for the Philosophy of Science*, 47, 579-607.

Horgan, T. (1997). Kim on Mental Causation and Causal Exclusion. *Philosophical Perspectives*, 11, 165-184.

Hüttemann, A. & Terzidis, O. (2000). Emergence in physics. *International Studies in the Philosophy of Science*, 14, 268-281.

Jackson, F. & Pettit, P. (1990). Program Explanation: A General Perspective. *Analysis*, 50, 107-117.

Kim, J. (1992a). Multiple Realization and the Metaphysics of Reduction. *Philosophy and Phenomenological Research*, 52, 1-26.

Kim, J. (1992b). Downward causation in emergentism and non-reductive physicalism. In Beckermann, A., Flohr, H. & Kim, J. (eds.) (1992). *Emergence or Reduction?* Berlin, New York: de Gruyter, pp. 119-138.

Kim, J. (1993a). *Supervenience and Mind*. Cambridge: Cambridge University Press.

Kim, J. (1993b). Can Supervenience and 'Non-Strict Laws' Save Anomalous Monism? In Heil, J. & Mele, A. (eds.), *Mental Causation*. Oxford: Oxford University Press, pp. 19-26.

Kim, J. (1996). *Philosophy of Mind*. Boulder, Oxford: Westview Press.

Kim, J. (1997a). Does the Problem of Mental Causation Generalize? *Proceedings of the Aristotelian Society*, 97, 281-297.

Kim, J. (1997b). Supervenience, emergence, and realization in the philosophy of mind. In Carrier, M. & Machamer, P. K. (eds.), *Mindscapes: Philosophy, Science, and the Mind*. Konstanz, Pittsburgh: Universitätsverlag Konstanz, University of Pittsburgh Press, pp. 271-293.

Kim, J. (1999a). Supervenient Properties and Micro-Based Properties: A Reply to Noordhof. *Proceedings of the Aristotelian Society*, 99, 115-118.

Kim, J. (1999b). Making sense of emergence. *Philosophical Studies*, 95, 3-36.

Kim, J. (2001). An Interview With Jaegwon Kim. (By Robert Howell, October 2000.) In Ephiloosopher. Online at <http://24.86.132.253/010101/symposia/symposia.htm>

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

Llinás, R. (2001). *I of the Vortex: From Neurons to Self*. Cambridge, London: MIT Press.

Lewis, D. (1973/1978/1986). Causation; Postscripts to Causation. In *Philosophical Papers, Vol. II*. New York, Oxford: Oxford University Press, pp. 159-213.

Lewis, D. (1980). Mad Pain and Martian Pain. In *Philosophical Papers, Vol. I*. New York, Oxford: Oxford University Press 1983, pp. 122-130.

Lewis, D. (1986). Causal Explanation. In *Philosophical Papers, Vol. II*. New York, Oxford: Oxford University Press, pp. 214-240.

Loewer, B. (2001). Book Review of *Mind in a Physical World* by Jaegwon Kim. *Journal of Philosophy*, 98, 315-324. Online at <http://fas-philosophy.rutgers.edu/~loewer/kimrev.pdf>

Loewer, B. (forthcoming). Review of Mind in a Physical World by Jaegwon Kim. *Philosophy and Phenomenological Research*. Online at <http://fas-philosophy.rutgers.edu/~loewer/kimppr1.pdf>

Lowe, J. (2001). Mind in a Physical World: An Essay on the Mind-Body Problem and Mental Causation. By Jaegwon Kim. *The American Journal of Psychology*, 114, 303-308.

Mandik, P. (2001). Mental representation and the subjectivity of consciousness. *Philosophical Psychology*, 14, 178-202.

Marras, A. (1998). Kim's Principle of Explanatory Exclusion. *Australasian Journal of Philosophy*, 76, 439-451.

Marras, A. (2000). Critical Notice of Mind in a Physical World by Jaegwon Kim. *Canadian Journal of Philosophy*, 30, 137-160. Online at http://www.swif.uniba.it/lei/mind/forums/003_0002.htm

Meijers, A. W. M. (2000). Physical and Mental? Reply to John Searle. *International Journal of Philosophical Studies*, 8, 179-183.

Menzies, P. (forthcoming). The Causal Efficacy Of Mental States. Online at <http://www.phil.mq.edu.au/staff/pmenzies/Mental%20Causation.pdf>

Meyering, T. C. (2000). Physicalism and Downward Causation in Psychology and Special Sciences. *Inquiry*, 43, 181-202.

Nagel, E. (1961). *The Structure of Science*. New York: Harcourt, Brace.

Nagel, T. (1974). What is like to be a bat? *Philosophical Review*, 83, 435-450.

Nagel, T. (1980). The Limits of Objectivity. In McMurrin S. M. (ed.), *The Tanner Lecture on Human Values, Vol. I*. Cambridge: Cambridge University Press, pp. 75-139.

Newman, D. (2000). Jaegwon Kim. Mind in a Physical World. *Journal of Consciousness Studies*, 7(10), 89-90.

Nordhof, P. (1999). Micro-Based Properties and the Supervenience Argument: A Response to Kim. *Proceedings of the Aristotelian Society*, 99, 109-114.

Polger, T. W. (2000). Review of Jaegwon Kim "Mind in a Physical World": An Essay on the Mind-Body Problem and Mental Causation. *Philosophical Psychology*, 13, 135-139.

Popper, K. R. & Eccles, J. C. (1977). *The Self and Its Brain*. Berlin: Springer.

Rockwell, T. A Defense of Emergent Downward Causation. Online at <http://www.california.com/~mcmf/causeweb.html>

Savellos, E. E. & Yalçın, Ü. D. (1995). *Supervenience. New Essays*. New York: Cambridge University Press.

Searle, J. (1995). Consciousness, the Brain and the Connection Principle. A Reply. *Philosophy and Phenomenological Research*, 55, 217-232.

Searle, J. (2000). Mental Causation, Conscious and Unconscious: A Reply to Anthonie Meijers. *International Journal of Philosophical Studies*, 8, 171-177.

Shear, J. (ed.) (1998). *Explaining consciousness: The Hard Problem*. Imprint Academic: Thorverton.

Stephan, A. (1999). *Emergenz. Von der Unvorhersagbarkeit zur Selbstorganisation*. Dresden, Munich: Dresden University Press.

Sturgeon, S. (1998). Physicalism and Overdetermination. *Mind*, 107, 411-432.

Tye, M. (1995). *Ten Problems of Consciousness*. Cambridge: MIT Press.

Vaas, R. (1995). Reduktionismus und Emergenz. In *Die mechanische und die organische Natur. Beiträge zum Naturverständnis. Konzepteheft Nr. 45 des SFB 230*. Stuttgart, Tübingen, pp. 102-161.

Vaas, R. (1999). Why Neural Correlates Of Consciousness Are Fine, But Not Enough. *Anthropology & Philosophy*, 3, 121-141. Online at <http://www.swif.uniba.it/lei/mind/texts/t0000009.htm>

Vaas, R. (2000a). Review of John Haugeland: Having Thought. *Erkenntnis*, 52, 139-147.

Vaas, R. (2000b). Evolving language, I-consciousness and free will. In Dessalles, J.-L. & Ghadakpour, L. (eds.), *Evolution of Language. Proceedings of the 3rd International Conference, 3-6 April 2000*. Paris: Ecole Nationale Supérieure des Telecommunications, pp. 230-235. Online at http://www.infres.enst.fr/confs/evolang/actes/_actes73.html

Vaas, R. (2001a). It Binds, Therefore I Am! Review of Rodolfo Llinás' 'I of the Vortex'. *Journal of Consciousness Studies*, 8(4), 85-88. Online at <http://www.imprint.co.uk/pdf/1200.PDF>

Vaas, R. (2001b). Willensfreiheit. In *Lexikon der Neurowissenschaft, Vol. 3*. Heidelberg, Berlin: Spektrum Akademischer Verlag, pp. 450-459.

Vaas, R. (2001c). Why Quantum Correlates Of Consciousness Are Fine, But Not Enough. *Informação e Cognição*, 3. Online at <http://www.marilia.unesp.br/atividades/extensao/revista/v3/artigo4.html>

Vaas, R. (2002a). Consciousness and its place in nature. *Journal of Consciousness Studies*, 9(2), 69-78. Online at <http://www.imprint.co.uk/pdf/1256.PDF>

Vaas, R. (2002b). Der Streit um die Willensfreiheit. Die Grenzen unserer Autonomie. *Universitas*, 57, 598-612 & 807-819.

Vaas, R. (2002c). Der Intelligenzsprung. *bild der wissenschaft*, 8, 30-39.

Van Gulick, R. (2001). Reduction, Emergence and Other Recent Options on the Mind-Body Problem: A Philosophic Overview. *Journal of Consciousness Studies*, 8(9-10), 1-34.

Whiten, A. & Byrne, R. (eds.) (1997). *Machiavellian Intelligence II*. Cambridge: Cambridge University Press.

Williams, M. (2000). Mind in a Physical World - Jaegwon Kim. *International Philosophical Quarterly*, 40, 377-8.

Witmer, D. G. (2000). Locating the Overdetermination Problem. *The British Journal for the Philosophy of Science*, 51, 273-286.