Michael Schmitz*,†

The Microstructure View of the Brain-Consciousness Relation

1 Outline

How can consciousness, how can the mind be causally efficacious in a world which seems—in some sense—to be thoroughly governed by physical causality? Mental causation has been a nagging problem in philosophy since the beginning of the modern age, when, inspired by the rise of physics, a metaphysical picture became dominant according to which the manifest macrophysical world of rocks, trees, colors, sounds etc. could be eliminated in favor of, or identified with, the microconstituents of these entities and their basic physical properties, plus their effects on human or animal minds. Against the background of this ontology, the argument from causal closure, or the causal completeness of physics, exerts strong pressure to also identify consciousness with microphysical entities—or even to eliminate it in favor of the latter—the only other options apparently being either the denial of the causal closure of the physical level, epiphenomenalism about the mind, or the view that its physical effects are generally overdetermined. In this paper, however, I want to introduce what I call the "microstructure view" (MV) of the brain-consciousness relation, and I want to try to make plausible that the problem of mental causation can also be solved, or perhaps rather dissolved, on the basis of this account. On the MV, the minimal neuronal correlates of consciousness—of the global state of consciousness, or specific states of consciousness such as pain—are not identical with these states, but rather constitute their microstructure, or, as I shall also say, equivalently, compose them. For example, if a certain pattern of neuronal activity in the prefrontal cortex is the minimal neuronal correlate of a certain kind of intention— that is, if it is both necessary

* Universität Konstanz. E-Mail: Michael.Schmitz@uni-konstanz.de.

† For helpful comments and criticisms I thank audiences in Konstanz, Berlin and London and especially Tim Crane, Christoph Fehige, André Fuhrmann, Gottfried Seebaß, Scott Sturgeon, David Papineau and Bernhard Thöle.

and sufficient for the presence of this intention—I shall also say that it is the neuronal microstructure of this intention and that it composes it.

The reason that there is also a sense in which the MV dissolves the problem of mental causation is that it abandons the metaphysical background picture against which it is commonly posed. It is essential to the MV that an identification (or elimination) is rejected not only for consciousness, but also for the manifest macrophysical world—and of course also for unconscious mentality, if such there be. Mental causation can only be adequately understood as one kind of macro level causation among others. So the microstructure relation is supposed to hold not only between states of consciousness and their neuronal correlates, but also between, say, plants and the cells that compose them and the molecules, atoms and ultimately subatomic entities that compose these in turn.

The guiding idea of the account of mental causation is that there is no competition for causal powers between macro phenomena and the micro level phenomena composing them and thus also no conflict between the causal closure at the microphysical level and the reality of mental causation. Micro level causation does not preempt macro level causation; it rather explains it.

The outline of the paper is as follows. I will first present and analyze the argument from causal closure in more detail. In the course of this analysis, I will briefly say why we should not choose one of the options of denying the causal completeness of physics, of epiphenomenalism, overdetermination or the identity theory, and I will explain why we also do not need to choose between them. The brief critique of the identity theory will naturally lead on to a more thorough exposition of the MV, culminating in the promised account of mental causation.

2 The Problem of Mental Causation

Consider the following four propositions:

(1) Mental entities (states, events, properties, facts etc.) have physical effects.

(2) The physical domain is causally closed: each physical effect has a sufficient physical cause.

(3) Mental entities are real and distinct from physical ones.

(4) The effects of mental entities are not generally overdetermined.

Now the familiar problem is that these propositions cannot be jointly true. Any three of them will entail the negation of the fourth. So if the mind is causally efficacious in the physical domain (1), and this domain is causally closed (2), but mental entities are real and distinct from physical ones (3), then the effects of consciousness entities must be overdetermined ($\sim$ 4). Or if the effects of mental entities are not generally overdetermined (4), and mental entities are real and distinct from physical ones (2), and the physical domain is causally closed (3), then mental entities cannot be causally efficacious in the physical domain ($\sim$ 1). Likewise one could argue for the falsity of the principle of causal closure and thus for ($\sim$ 2) on the basis of (1), (3) and (4). The most important and influential line of argument, however, has been to argue for the identity theory and thus for ($\sim$ 3) on the basis of (1), (2) and (4).

I shall now first argue that neither epiphenomenalism, overdetermination nor the identity theory are plausible options. My brief discussions are not intended to do justice to all the issues raised by these positions. This would obviously be impossible in a short paper like this. Rather the point is to draw some lessons from the failure of these positions that will prepare the positive solution to be presented later and to provide some motivation for accepting that solution, or at least for taking it seriously.

3 Epiphenomenalism

Discussions of epiphenomenalism sometimes begin by introducing a distinction between those properties of an event that are causally relevant and those which are not. For example, the fact that the whiskey I had contained alcohol causally explains the fact that the three shots I had made me drunk, while the fact that the whiskey was made in Tennessee does not (Bieri 1992). There are two points worth noting about this for present purposes. First, such examples only show that an entity may have properties which are *causally irrelevant relative to certain effects*. But they fail to establish the sort of *absolute causal irrelevance* which the epiphenomenalist asserts for consciousness. The property of having been made in Tennessee may be irrelevant to the causal explanation of my drunkenness, but it certainly is not causally irrelevant per se. For example, the fact that it was made in Tennessee may have played a causal role in my decision to choose this particular kind of whiskey. So examples of this kind cannot be used to support the notion of an absolutely causally irrelevant property. Second, note that the fact that the property of being composed of alcohol caused the drunkenness does not show that the whiskey, or, more strictly, the event of my drinking whiskey, is causally irrelevant as such. We would not ordinarily conclude that whiskey is incapable of causing drunkenness. The natural thing to say here would rather be that the fact that whiskey

contains alcohol *explains* why and how it can cause drunkenness. I shall later argue that we should think about mental causation in a similar way.

Epiphenomenalism is so implausible that it is not necessary to argue against it in the present context. Only somebody despairing of solving the problem of mental causation will be tempted by it. But we shall see that there is no reason for such despair.

4 Overdetermination

Overdetermination would occur if there were two or more independent and yet equally sufficient causes of the same effect. It is often suggested that there are actual instances of overdetermination. If this were true, it would seem to make the overdetermination option more acceptable. It would still be odd that the physical effects of mental causes should always be overdetermined, but this might seem tolerable. After all, embracing overdetermination would solve the problem of mental causation! I want to make plausible though that the alleged examples of overdetermination are unconvincing. The critique of these examples will show that overdetermination is not a genuine option and will also teach us an important lesson about causal explanation.

Consider a standard example, like a man shot with several bullets. Each of these bullets, we assume, would have been sufficient by itself to kill him. Does it not follow that there are several independent and yet equally sufficient causes of the same effect which is thus overdetermined? It seems to me it does not. From the fact that each bullet *would have* killed the man even in the absence of the others, it does not follow that some *actual* event was overdetermined. When we track the causal paths of the individual bullets through the body, it seems clear that at some level the bullets always cause different things. This should be obvious from their different spatio-temporal paths through the body. It is a consequence of this that in some cases it will be clear that one bullet was the cause. For example, it might enter the brain and instantly terminate all life-functions. This bullet will clearly be the cause of the death, notwithstanding the fact that, if it had not killed, the man would have died minutes, hours or weeks later, from the damage done by another bullet. That fact as such is just irrelevant to the assessment of what actually caused the death. This kind of case is often referred to as one of causal 'preemption.' The actual cause has preempted a potential different cause. In other kinds of cases, it seems more appropriate to say that different causal paths converge, so that two or more bullets jointly cause the death. This will be so if the further bullets cause the death to happen differently than it would have, had they not been present. For example, several bullets may enter the heart simultaneously,

causing greater damage than each of them would have individually and an earlier and less painful death.

It is sometimes objected against this way of thinking that it makes too many properties of the effect event essential, which would thus become very 'fragile.' But this reasoning is fallacious. Of course the death might have occurred in countless other ways. But we causally explain the actual death, not some counterfactual death. However a distinction between essential and accidental properties of a particular event is drawn, if at all: causal explanations of events can and will also explain their accidental properties. These are also part of the way the event is. A related move (Funkhouser 2002) is to introduce a notion of an event like a death simpliciter, with no further specification of the manner of death. The idea then is that the death simpliciter is overdetermined in any case. But this move is hardly convincing either. Obviously there is no death or other event which does not happen in a specific way, and one can hardly prove there is overdetermination simply by refusing to specify the relevant effects further.

We can draw a lesson from the preceding discussion about when causes may be said to compete with one another. The bullets can be said to causally compete with one another in the sense that there may be a meaningful question as to which actually caused a given phenomenon. As we have seen, in some cases it will be correct to say that one rather than the other caused a certain death. Mental and physical causes also sometimes compete in this way. For example, we often ask whether the cause of certain disease symptoms is psychological or physiological, and in some cases an unambiguous answer to this question is possible. I want to suggest though that this is only possible when the putative causes operate on the same level. Causes at different levels do not compete with one another. Intentions, or brute physical forces may both cause my arm to move and so it makes sense to ask, in a given situation, which kind of cause was responsible for the arm movement. But we cannot ask in the same way whether the intention or its neurophysiological correlate caused the arm to go up. The reason is that the correlate is not even a candidate for being the cause of the arm movement. It causes things at the physiological level, and the arm is not at that level, though it has parts which are. Or so I shall argue more extensively later. To see that mental events and their physiological correlates do not causally compete with one another because they operate at different levels will be the key to solving, respectively dissolving the problem of mental causation.

5 Identity Theory

It is not my goal in this paper to provide a decisive refutation of the

identity theory. This would be a task for one or more separate papers.[1] However, it will be useful to take a brief look at the central difficulty of the identity theory, focusing on its type identity version. This difficulty can be brought out in several different ways, but the following is perhaps the most straightforward. The identity theorist claims that a state of consciousness is identical with its minimal neuronal correlate—or, for example, water with the H_2O molecules it is composed of. But it seems hard to deny that knowing that one, for example, is afraid of the dark is different from knowing that one is in neuronal state X, where we will take "X" to abbreviate a specification of the minimal neuronal correlate of being afraid of the dark. Obviously, one can know one without knowing the other. Moreover, this difference seems to be a cognitively significant difference in the contents of the relevant knowledge—as opposed to a mere difference in the vehicles of representation. That is, it is not like the difference between a German speaker saying to himself "Ich habe Angst vor der Dunkelheit" and an English speaker saying to himself "I am afraid of the dark."

The identity theorist will now say something like that these different states provide different *modes of access* to the same entity.[2] He may liken this to the fact that, for example, the same tree can be seen from different perspectives. He may also appeal to a Fregean notion of sense, according to which words or concepts may differ in sense while still referring to the same entity. This strategy can certainly be made to sound plausible. Ultimately, however, it is difficult to understand. For in the examples commonly given, the difference in perspective or sense always involves different properties of what is accessed, such that a difference with regard to the mode of access would seem to be inextricably linked to objective ontological differences in what is accessed. So, for example, when I see a tree from different perspectives, I ipso facto see it as standing in different spatial relations to me, and I will also access different inherent properties of the tree when seeing it from different sides. Likewise, the expressions "animals with a heart" and "animals with a kidney"—to use a time-worn example for the sense/reference or intension/extension distinction—can also only pick out the same class of animals in cognitively significant different ways because they make reference to different properties of these animals. So if these cases are taken as a model—and what else could we use as such—it seems the most the identity theorist might achieve is to replace a difference of things or events through a difference in the properties of these events or things. This would amount to what is usually referred to as a token identity

[1] I have given a much more extensive critique of the identity theory in my yet unpublished dissertation. Other recent critiques of the identity theory similar in spirit to mine include McGinn (2001) and Horgan & Tienson (2001).

[2] Very many authors have defended the identity theory in this way. Compare, for example, McLaughlin (2001), Pauen (1999) and Perry (2001).

thesis. But it will nowadays be generally accepted that this is not enough to solve the problem of mental causation; irreducible mental properties are just as offensive from the physicalist point of view as irreducible mental events, and to account for the causal relevance of the former is as much a problem as to account for the causal relevance of the latter.

The point can also be made by saying that the examples do not make intelligible how it would be possible for the same *fact* to be known or accessed in different ways. They only show that the same thing or event might be accessed in different ways by being accessed as the bearer of different properties and thus by apprehending different facts about it. This cannot be a model for how it might be possible to know the same fact in different ways. But such a model is precisely what we need in order to understand how knowing that one is afraid and knowing that one is in brain state X might be knowing the same thing in different ways, because obviously it is knowledge of facts that we are dealing with here. Moreover, we have already pointed out that it will not help to appeal to additional distinct properties, simply because that will subvert the purposes of the identity theorist. So it seems that the identity theorist is stuck with an unexplained notion of what I will call "merely subjectively different" modes of access or ways of knowing. To call these "merely subjective" is to mark the fact that there is supposed to be a cognitively significant difference between the relevant intentional states which does not correspond to any objective difference in the (putative) objects of these states, that is, the states of affairs they are directed at. I have argued that we have no model in ordinary thought for this decidedly metaphysical idea; nor do I know how it can be made intelligible in some other way. It thus seems best to avoid it.

6 The Principle of Causal Closure

Flatly denying the principle of causal closure also does not seem to be an attractive option. There is a lot of empirical evidence for at least some version of the principle. Throughout the history of science, it has been shown for a wide range of initially apparently recalcitrant phenomena that they can be explained in terms of basic physical entities and forces.[3] This has been shown for chemical and biological phenomena, and there seems to be no good empirical reason to think that mental phenomena will prove to be an exception. As far as I know, there are no solid scientific data supporting the view that consciousness, to put it crudely, can do things that cannot be explained physiologically and ultimately microphysically.

[3] For an account of the relevant history of science, see the appendix of Papineau (2002).

Are we thus in deep trouble? Fortunately, though, there is still room for disagreement how the principle should be formulated and understood and what it entails. I want to argue now that there is no reading in which "physical" in (1) and (2) is univocal and both come out as plausibly true.[4]

The causal powers we commonly attribute to the mind are causal powers at the common sense *macrophysical* level. They are powers to move the human or animal body and other macroscopic entities by moving that body. By contrast, causal closure only obtains at the microphysical level, presumably that of quantum mechanics, or an even lower level. Only at that level can everything be completely causally explained in terms of its level alone. This is so because only at that level, since all entities are composed of microphysical entities, there is literally no external causal influence. That is not true for any of the levels above the microphysical level. For example, on the biological level there is obviously continuous external causal influence: organisms crucially depend on their non-biological environment, on the causal powers of the sun, the atmosphere etc. So the principle of causal closure is only true if "physical" is taken to mean "microphysical." But then it does not exclude causation above the microphysical level, neither mental nor other kinds of macro causation.

It is important to emphasize that, even after the mistake of equivocating on different senses of "physical" is corrected, the principle of causal closure, while not forcing us to embracing either epiphenomenalism, overdetermination or the identity theory, still does put an important constraint on our account of mental causation. I put this above by saying that it must be possible to explain the causal powers of the mind with reference to the physiological and, ultimately, the microphysical level. Mental causation and macro causation generally must be micro-explainable. Another way of saying this is that the causal powers of the mind cannot be free-floating with regard to the levels below it. The most popular way of expressing this, however, has been to say that the causal powers of the mind and the mind generally must be reducible. The success of the physical sciences makes it very likely, perhaps even virtually certain, that there are no irreducible mental causal powers. The concept of reduction is highly problematic, though, because it is usually understood in such a way that two ideas are rolled into one, namely that macro level entities are micro-explainable, most likely by means of biconditional correlation laws, and that they can be identified with micro level entities. But in this way the possibility of a position like the MV, which is reductionist in the first sense, but not in the second, is systematically concealed. So let me now set out the MV in more detail.

[4] For this diagnosis of a fallacy of equivocation I am indebted to Sturgeon (2000), whose version of a composition account has been a major source of inspiration for this paper.

7 The Microstructure View

The MV thus accepts the empirical assumptions of the type identity theory, but replaces its metaphysics through one based on the notion of microstructure (composition) and does so across the board, not only for consciousness, but also for the manifest macrophysical world. The central empirical assumption is that there are biconditional correlation laws linking macro and micro level entities. Note that these laws may be quite specific. Such a law might state, for example, that a certain kind of throbbing human ear pain will occur if and only if a certain neuronal condition obtains. This system of events at the neuronal level will be the microstructure of the pain, what composes it at the neuronal level. The neuronal level in turn will be composed of events at lower levels, down to the lowest microphysical level if such there be. How far science will actually progress in discovering laws of this kind is a question which cannot be answered from the armchair. It seems likely though that some generalization will be so specific that uncovering them would neither be worth the trouble nor even possible, given human epistemic limitations. The MV assumes that it is still reasonable to maintain that each macro level type has its own characteristic type of microstructure, its micro level signature, as it were, even if we may be unable to precisely delineate it in each case. This also means that supervenience will hold: there will be no macro level differences without corresponding micro level differences.[5]

The ontological difference between MV and type identity theories is reflected in the fact that the former, but not the latter, will think of the correlation laws as genuine laws of nature—even though they are different in kind from intra-level laws. The law statements will employ logically independent concepts with distinct cognitive significance, and the discussion of the identity theory was designed to make plausible that if such concepts refer, they must refer to different entities, on pain of avoiding the unpalatable notion of merely subjectively different modes of access. It is worth emphasizing that the MV does not only reject a posteriori identities of this kind, but also priori identities (or eliminations) based on conceptual analysis. The MV is thus also incompatible with any account trying to analyze the notion of a given whole solely in terms of its parts, for example through purely formal operations such as mereological operations. The crucial point is that there are logically independent macro level concepts, and that if these concepts succeed in referring to something, it must accordingly be to ontologically distinct entities. So if Martians came to earth already possessing a more or less complete microphysical understanding of

[5] I cannot here address issues related to externalism and the attendant question whether supervenience will be local or global.

what is going on here, they would still need to acquire plenty of new, logically independent concepts of plants, of animals, people and their mental states, of fashion trends and government institutions, and these concepts would give them access to corresponding new entities so far unknown to them.

It is a consequence of this account that macro and micro level entities cannot be distinguished in terms of their size alone. Micro entities can get as big as you like. The spacetime content of the entire universe described solely at the level of quantum mechanics would still be a micro level entity. An aggregate of neurons picked out only in neuronal level terms is just a complex state of affairs at the neuronal level, certain neurons standing in neuronal level relations to one another. What matters is the size of the entities picked out by the *primitive* concepts of a given level such as "quark," "neuron," or "skirt."

The composition relation prototypically relates things. However, it is also common to apply compositional concepts to events. We will then ordinarily say that one event *consists* in others. For example, the football World Cup consists in various matches, which in turn consist of various plays and actions. An election consists in a multitude of events involving the voters and various officials. In essentially the same way I will also talk about composition relations between levels, like the composition relation between a conscious intention in consciousness and its neuronal level correlate, which will likely be some system of neuronal events, like, just to give a possible example, an assembly of neurons firing in synchrony at a certain frequency.

8 The Microstructure View and Mental Causation

The topic of events naturally leads to the MV account of mental causation and macro causation generally, which, following the dominant tradition of thought, I will take to be a relation between events. I will not say more about what I think causation is. In particular, I intend to be neutral with regard to the issue of counterfactual vs. regularity vs. mechanism accounts of causation. This appears appropriate, as the most important point to be made in this paper is that macro and micro level causation do not compete with one another. It seems to me this claim should come out as true no matter what one's account of causality is. It should be a condition of adequacy on a satisfactory account of causality.

The key to the MV account of macro causation is to see that causal relations are relations within a given level, that relations between levels are compositional and lawlike rather than causal, and that the compositional structure of higher level entities explains their causal interactions rather than draining their causal powers away. Given this, we will be able to say,

for example, that the intention causes the arm to go up without violating a reasonable version of the principle of causal closure, because the macro level cause will not cause anything on the micro level. Micro level causation at the neuronal and lower levels will go on, unperturbed as it were, only connected to the macro level through compositional and lawlike relations, which, because they are different in kind from causal relations, do not overdetermine their relata, while still explaining macro level causality. The MV account is thus a version of what is often referred to as a dual explananda account of mental causation: there is no conflict with microphysical causal explanations because these explain different phenomena. I will now defend the principles behind this account.

Some authors have claimed that the microstructure causes the macro entity ('upward causation'), some that the macro entity acts causally on its micro level correlate ('downward causation'). Both these views sit uncomfortably with the dominant conception of causality and for the same reason. The macro entity and its microstructure are neither spatially nor temporally distinct. The microstructure fills the same spacetime content as the macro level entity. So it seems odd to say that the brain causes the presence of its consciousness level correlate (Searle 1992) or that a wheel rolling downhill causes the movements of the micro entities it is composed of (Sperry 1969). Intuitively, no causal interaction is taking place here. The wheel is not pushing its constituents along. We may still say that the fact that the wheel rolls downhill explains the fact that its microconstituents move in a certain way. It explains this because the microconstituents compose the wheel. That is why they have to go along for the ride. So this is explanation through composition rather than causation. But this is of course a rather trivial sort of explanation. The more interesting direction of explanation is the opposite one, where we explain the motion of the wheel in terms of the causal relations between the micro entities composing it, and the causal interactions between the system of micro entities and its environment. Likewise, it seems clear that the presence of a neuronal correlate of consciousness does not cause the conscious state in any ordinary sense of causation. It is not that the firing neurons trigger some process which results in the relevant state of consciousness. And while there is a law present here, this is generally not thought to be sufficient for the presence of causality. Moreover, a biconditional law like a correlation law is particularly unsuited to underwrite the claim that the relation is causal, as its existence gives us no reason to think that the brain state causes the state of consciousness rather than conversely.

Now, of course Searle and Sperry are aware that their understanding of causation is a non-standard one. Suffice it to note that for present purposes it does not really matter in the end whether the inter-level relations should be called "causal" or not. We might of course decide to apply the term "causality," for example, to relations of mutual dependence between

synchronous entities. (There is certainly an element of natural necessity here, which provides at least some motivation for this extension of the concept of causality.) The crucial point is that these relations are different in kind from the prototypical instances of causality. This is true whether we think of these relations in terms of different kinds of causal relations, or of causal vs. non-causal relations. In accordance with the standard notion of causality, I prefer to think of them as being non-causal though.

Because the micro-macro relations between minimal correlates and the causal relations within levels are in any case different in kind, the fact that one can explain some micro and macro effects both through their causes at the same level and, by subsuming them under a correlation law, through their micro or macro correlates, does not, I think, constitute overdetermination in any objectionable sense. Obviously, these cases are very different from the putative examples of overdetermination discussed above, which would involve spatiotemporally distinct, independent causes at the same level. If somebody still insisted that there can only be one explanation for a given phenomenon of whatever kind, I do not know what the justification for such a principle would be. There seems to be no a priori justification for it. Moreover, it would most likely make any satisfactory (dis)solution of the problem of mental, or, more generally, macro causation, impossible. But we are definitely much more certain of the fact of macro causation than of the truth of any such principle.

Why do micro level causal interactions *explain* macro level ones? Intuitively, it seems obvious that they do. When we track the causal pathways from the correlates of the intention to the physiological events in the arm that cause the micro-correlate of the muscle contraction, we *understand* much better how an intention can cause a bodily movement. Slightly more technically, we can also say that the micro-macro relations are explanatory because they *integrate* our knowledge at the mental and common sense macrophysical level with our physiological theories. Furthermore, the discovery of the microstructure of the mind also makes it possible to integrate our physiological theories of mind and body. We will see in which respects the physiology of the mind is continuous with that of the body and in which ways it is special. The result of all this is a more *unified* and thus more explanatory picture of the world. Finally, it seems to me there is no clear sense to the idea that micro level entities take away causal powers from their macro correlates. After all, that macro entities have the microstructures that they do, is part of what they are, their empirically discovered nature.

9 Diagonal Causation and Overdetermination

A critic who accepts the argument so far might now object as follows: "I

will grant that the synchronous micro-macro relations are not causal. The notion of causation requires a temporal difference between cause and effect. But what about the neuronal correlate of the intention to raise the arm? That surely looks like a cause of the raising of the arm. In this case, the relevant events are temporally and spatially distinct. So isn't this an instance of overdetermination after all? It seems that both the macro level intention and its micro correlate cause the raising of the arm, which is thus overdetermined."

The critic appeals to "diagonal upward causation," a causal relation between a prior micro level and a posterior macro level event. Correspondingly, "diagonal downward causation" would occur when a temporally prior macro level event causes a micro level event, as when the intention causes the physiological correlate of the arm going up. Now, should we recognize diagonal causation, and does it lead to overdetermination?

To deny that diagonal relations are causal would needlessly put us at odds with common sense. We would be forced to deny, for example, that a virus—or, more strictly, the entry of a virus into a human body—could cause the outbreak of a disease. But we do not need to deny this, because diagonal causation does not lead to overdetermination. I have already argued that the fact that an entity can be explained both through its intra-level causal and its inter-level compositional relations does not lead to overdetermination. I now want to extend that argument by showing that, when all facts about intra-level and inter-level relations are fixed, there is no further causal work to do. Diagonal relations are an automatic fallout from these relations, and talk about "diagonal causation" thus does not introduce any additional, independent causal factors and does not lead to overdetermination.

To see this, let us take a closer look at the virus example. The example is an instance of diagonal causation because the virus belongs to a micro level relative to the disease, which is a macro level condition affecting various organs and the well-being of the whole person. The microstructures of the affected parts of the body (or mind) will change in ways characteristic of the disease. These changed structures are its microstructure. The evidence for the diagonal causal claim consists in a strong correlation between the presence of the virus and the outbreak of the disease. But now suppose it turns out that the virus is not actually responsible for the microstructural changes characteristic of the disease. These are, we may assume, rather caused by prions thriving in the same environment as the virus. The point now is that under these circumstances the virus also has no claim to be the diagonal cause of the disease. It would not make sense to accept the facts as described in this scenario and yet to insist that the virus causes the disease. The virus cannot cause the disease except by, or independently of, causing its microstructure. Analogous remarks apply to diagonal downward causation. The intention to raise the arm has no claim

to be the cause of the physiological happenings which compose the raising of the arm if it does not actually cause that raising. If some other macro level phenomenon caused the raising, that phenomenon will also be the diagonal cause of the raising's microstructure.

Consideration of these examples shows that there are no further, additional facts about whether diagonal causal relations obtain, facts independent of facts about prototypically causal relations within levels and facts about relations between levels like compositional facts. For if there were such facts, it would have to be possible, for example, for the virus to diagonally cause the disease without causing its microstructure. Since that does not appear to be possible, I am drawn to the conclusion that talk about diagonal causation is a sort of abbreviation for talk about certain constellations of causal intra-level and compositional inter-level facts. Diagonal causal explanation is a thus a kind of hybrid explanation tracking both micro-macro and the prototypically causal intra-level relations. That is, to say that the correlate of the intention diagonally causes the arm to be raised is just a way of saying that it causes a micro level entity which composes this macro event. To say that the intention causes this micro level entity is just a way of saying that it causes some macro level entity which is composed by the micro entity. This is all there is to diagonal causation. There are no independent causal forces operating diagonally, and therefore there is also no overdetermination through diagonal causation.

10 Conclusion

To wrap things up, let me explain how the proposed account of mental causation depends on the abandonment of the received metaphysical picture with its elimination of the manifest macrophysical world. I have argued that we can make sense of mental causation if we clearly distinguish compositional inter-level and causal intra-level relations. We can see then that mental causation does not violate the causal closure of the microphysical level, but is rather explained by causal relations at lower levels. But suppose now we would accept the MV for consciousness only, while still trying to eliminate the manifest macrophysical world in favor of its microconstituents and their effects on the mind, that is, the mere appearance of this world. The simple reason why this cannot work is that there would be nothing left for the mind to cause—at least nothing physical. We still could not make sense of our natural conviction that the mind can also bring about changes in the physical world, because, since this world will be an exclusively microphysical world, the mind would either have to violate causal closure by perturbing the microphysical entities, overdetermine them, or be identical with some of them. So we would be back to our original predicament. The MV account thus

essentially depends on the rejection of the traditional metaphysics, and to the extent that it does, it dissolves rather than solves the problem, because one of the key assumptions defining the problem in its traditional form is simply abandoned. Given this initial dissolution, the problem can then, if the argument of this paper is correct, be solved in a fairly simple way. But the fact that the dissolution makes the solution available is also a powerful argument in favor of it, especially in view of the fact that all other options appear to be extremely implausible.

REFERENCES

Bieri, P. (1992). Trying out epiphenomenalism. *Erkenntnis* 36, 289-309.

Funkhouser, E. (2002). Three varieties of causal overdetermination. *Pacific Philosophical Quarterly* 83, 335-351.

Horgan, T. & Tienson, J. (2001). Deconstructing new wave materialism, *Physicalism and its Discontents*, ed. by C. Gillet & B. Loewer. Cambridge: Cambridge University Press.

McGinn, C. (2001). How not to solve the mind-body problem, *Physicalism and its Discontents*, ed. by C. Gillet & B. Loewer. Cambridge: Cambridge University Press.

McLaughlin, B.P. (2001). In defense of new wave materialism: A response to Horgan and Tienson, *Physicalism and its Discontents*, ed. by C. Gillet & B. Loewer. Cambridge: Cambridge University Press.

Papineau, D. (2002). *Thinking about Consciousness*. Oxford: Oxford University Press.

Pauen, M. (1999). *Das Rätsel des Bewusstseins*. Paderborn: mentis.

Perry, J. (2001). *Knowledge, Possibility and Consciousness*. Cambridge, MA: MIT Press.

Schmitz, M. (2005). *Das Bewusstsein erklären*. Unveröffentlichte Dissertation, Universität Konstanz.

Searle, J. (1992). *The Rediscovery of the Mind*. Cambridge, MA: MIT Press.

Sperry, R. (1969). A modified concept of consciousness. *Psychological Review* 76, 532-536.

Sturgeon, S. (2000). *Matters of Mind*. London: Routledge.