Uwe Meixner

Physicalism, Dualism, and Intellectual Honesty

Mainly by offering a detailed criticism of distilled versions of the two main lines of argument for (ontological) physicalism, the causal-closure-line and the functionalist line, this paper aims to show that the rational support for physicalism (or materialism) is less strong than it is usually taken to be by its many adherents. In fact, physicalism is a metaphysical position, far exceeding what is vouched for by empirical science. Physicalism, therefore, does not have a greater rational claim on our credence than any other metaphysical position, for example, dualism (which position is given some argumentative support in the paper). Defending the view of – at least – metaphysical equality between dualism and physicalism also involves restating correctly and clearly what is often left unclear, or is explicated in a misleading way: what it is that physicalism and dualism basically amount to as ontological theses.

1 The purpose and the basis of this paper

By practically all of its proponents physicalism (or materialism; both terms are treated as synonyms in this paper) is held to be *the one and only rationally defensible world view*. I shall argue that this position is less motivated by good reasons than by the self-assurance of those who believe that they are on the winning team. In doing so, I will concentrate on the consideration of physicalism in the context of the philosophy of mind. In fact, the context of the philosophy of mind *alone* is the context to which I will be referring. Thus everything I say should be tacitly relativized to that context.

My first thesis, then, is this:

> *Thesis* 1: A comprehensive doctrine is called "physicalism" in the proper sense of the word only if it entails that all mental entities (i.e., substances, properties, events, states, objects, etc.) are physical.

The justification for *Thesis* 1 is easily given. Suppose a comprehensive doctrine is called "physicalism" in spite of the fact that it does not entail that all mental entities are physical. Hence this doctrine is compatible with

(psychophysical) *dualism*, according to which at least *some mental entities are not physical*. But a comprehensive doctrine that is compatible with dualism can certainly not be called "physicalism" in the proper sense of the word.

Though rather obvious, *Thesis* 1 does not appear to have been much heeded in recent times. Many philosophers have believed that all that physicalism, as a comprehensive doctrine, need amount to is the *supervenience* of the mental on the physical. Supervenience physicalism is often maintained to be a more plausible form of physicalism than any doctrine that is called "physicalism" in accordance with *Thesis* 1, according to which thesis physicalism, properly so-called, minimally amounts to what might be called the *General Identity Theory*: the doctrine that *everything mental is identical with something physical*. The question necessary to ask here is the following: Does the supervenience of the mental on the physical entail that all mental entities are physical? If *no*, then this position is compatible with dualism, and therefore a comprehensive doctrine of physicalism, *properly so called*, cannot be identified with it. The inevitability of this conclusion is frequently masked (to authors and readers alike) by claiming that supervenience materialism *does* entail that the mental is in a sufficiently strong sense *nothing over and above the physical*. But the main, the central question in the confrontation between dualism and materialism must surely be whether or not the mental is, at least in some instances, *something else* than the physical, not whether or not it is *something over and above the physical*, whatever this phrase may mean precisely if it is supposed to have a meaning that is different from the meaning of *something else than the physical*.[1] (If in some instances the mental proves to be *something else* than the physical, if, in other words, it turns out that some mental entities are not physical, then how could it be denied that dualism is true – even if the mental were, in some sense, nothing over and above the physical?)

If, on the contrary, the answer to the above question is *yes*, then why consider supervenience physicalism to be a more plausible form of physicalism than the General Identity Theory, which many consider to be *implausible*? Moreover, if supervenience materialism is supposed to be a more plausible form of materialism because it's "non-reductive,"[2] then, given the answer *yes* to the above question, one may well ask: what can be *non-reductive*, in an ontologically important sense (I am not talking here about linguistic or

[1] Presumably the idea is this: something might be *something else* than the physical without being *separable* from the physical, i.e., without being *something over and above* the physical. Yet, *being different* does seem to entail *being separable* (and vice versa), where "-able" expresses the appropriate notion of possibility: possible in the widest (objective) sense.

[2] The supervenience materialism of David Lewis, in contrast, is intended to be *reductive*; see Lewis (1994). Lewis notes (ibid., p. 414): "Yet thousands say that what's good about stating materialism in terms of supervenience is that this avoids reductionism!"

explanatory or any other non-ontological form of non-reductionism), about a materialism that entails that every mental entity is physical? It seems to me that, given the answer *yes* to the above question, this so-called non-reductive materialism must be in a straightforward sense *reductive* – for the following reason: if all mental entities *are physical* to begin with, then, trivially, they all *are reducible to something physical*.

It should be noted that even if the *A*s supervene on the *B*s in a very strong way, this does not ipso facto entail that the *A*s are *B*s, or that the *A*s are somehow reducible to the *B*s (which, when the process of reduction is over, would make them all *B*s after all). Consider the following example. *Lateral triangularities* are the properties such that each of them is nameable in the following manner: being a triangle with sides of the lengths *a*, *b*, and *c*. *Angular triangularities* are the properties such that each of them is nameable in the following manner: being a triangle with angles of the sizes *c*, *d*, and *e*. The angular triangularities supervene on the lateral triangularities in the subsequently described, very strong way (employing a concept of supervenience that, in the given context, is logically equivalent to a standardly used concept; cf. Kim (2001), p. 9, and Kim (1993), p. 65):

> For every angular triangularity *f*, it is *conceptually necessary* for every *x*: if *x* exemplifies *f*, then there is a lateral triangularity *g* such that *x* exemplifies *g* and such that it is *conceptually necessary* for all *y*: if *y* exemplifies *g*, then *y* exemplifies *f*.

Although the above assertion is *conceptually true* (and hence the operator of conceptual necessity can be prefixed to the assertion without changing the asserted content), it does not follow that the angular triangularities are lateral triangularities, or that the angular triangularities are somehow reducible to the lateral triangularities. On the contrary, with conceptual necessity, *no* angular triangularity is a lateral triangularity, because, with conceptual necessity, no angular triangularity has the same extension as any lateral triangularity. It is true that there is a one-to-one correspondence between angular triangularities and certain superdenumerably infinite sets of lateral triangularities. But one should resist the reduction of angular triangularities even to those sets, since angular triangularities are straightforward objects of cognition, while those sets certainly are not.

I need not dwell on the fact that the relationship that has, to date, been *established* to obtain between mental properties and physical properties comes *in no way* close to the relationship that obtains between the angular and the lateral triangularities. But even if that very relationship had been established to obtain also between the mental and the physical properties, that is, even if we had as a conceptual truth

> For every mental property *f*, it is *conceptually necessary* for every *x*: if *x* exemplifies *f*, then there is a physical property *g* such that *x* exemplifies *g* and such that it is *conceptually necessary* for all *y*: if *y* exemplifies *g*, then *y* exemplifies *f*,

it would not follow that mental properties are physical, or that mental properties are somehow reducible to physical ones. On the contrary, it could very well be true that not only some, but *all* mental properties are not physical. (To show this was the point of the above example.)

Eliminative materialists aside (who escape despair by trivializing physicalism: by denying the existence of mental entities[3]), I suspect that many of those who call themselves "physicalists" are quite aware that the prospects of the so-called *identity theories* (be they *token-* or *type-identity* theories) and of (ontological) *psychophysical reductionism* (which *ultimately* – namely, *after* the allegedly successful reductions – amounts to the very same thing as the identity theories) *are very dim indeed*,[4] and they have changed their claims accordingly. Nevertheless, they still uphold the names "physicalism" and "materialism" for designating their miscellaneous positions, and still despise "dualism," which they keep associating with religious obscurantism, alleging, with great emphasis, its anti-scientific nature.[5] They still treat "dualism" with a curious mixture of contempt and fear, never considering what it truly amounts to, as if it were not worthy of the attention of a rational person and, at the same time, something so terrible one had better not look at it, like the head of a Gorgon. All of this "enlightened" behaviour is merely good for one thing: to mask the fact that they who display it are no longer asserting anything that is in logical opposition to dualism, which is purely and simply the sober doctrine that *there is a non-physical side to the mental*, or more precisely: that *some mental entities are not physical*.[6] I would like to suggest that the time is ripe to return to intellectual honesty.

---

[3] According to predicate logic, "every M is φ" is trivially true if there are no Ms.

[4] At the end of this paper I hope to have shown them to be even dimmer; see Sections 4 and 5.

[5] For some documentation of this murkily ideological side of *some* physicalistically oriented thinkers, see Meixner (2004).

[6] There are stronger forms of dualism, of course. For example, the doctrine that not only some but *all* mental events are not physical, or the doctrine that subjects of experience are non-physical and can exist without a body, or even without anything physical existing. But there is no good reason to reserve the word "dualism" only for these logically stronger doctrines. Already *minimal* dualism – the doctrine that some mental entities are not physical – is *dualism* in the *proper* sense of the word, since already minimal dualism (if taken together with the obviously true claim that there are physical entities) maintains that there are *two* (nonempty) sides of (non-abstract) being: the physical and the non-physical, and that the mental is at least partly belonging to the latter.

It is easier to feel gratuitously offended at this than to escape the basic point I am making. Put in the most straightforward and non-polemical way I can think of, my basic point is this: Let anyone who believes that she is a physicalist and in opposition to dualism ask herself whether she believes that everything mental is physical. If the answer is "Yes," then she is a physicalist all right (hence no need to feel offended by my call for intellectual honesty!), but also an adherent of the frequently disavowed General Identity Theory (since "everything mental is physical" is logically equivalent with "everything mental is identical with something physical"). If however the answer is "No" – "No, I do not believe that everything mental is physical" –, then I indeed suggest that she is seriously deluding herself and others by calling herself a "physicalist."

Fittingly, my second thesis, complementing the first, is this:

> *Thesis* 2: If we use the word "physicalism" in the context of the philosophy of mind for a comprehensive philosophical doctrine, then we should use that word in its proper sense.

Why should we do that? Suppose in a philosophical discussion somebody upholds a comprehensive philosophical doctrine she calls "physicalism," and after a while it turns out that *her* physicalism is compatible with dualism, because it does not entail that all mental entities are physical. Is this a desirable situation? I should think not. In a philosophical discussion, we should not make it seem – not even prima facie – that we are claiming more than we really mean to claim, and of course we should also not make it seem that we are claiming less. For fulfilling this requirement, it is mandatory that we use our words in their proper senses (provided, of course, they have such).

In this paper, I shall use the word "physicalism" in its proper sense to designate a comprehensive philosophical doctrine. Hence (according to *Thesis* 1) physicalism entails that all mental entities are physical. In other words, any doctrine that does not entail that all mental entities are physical is *not* physicalism (in the proper sense of the word).

My third, and central, thesis is this:

> *Thesis* 3: Neither philosophical nor empirical considerations, nor their combination, suffice to render it probable, or even plausible, that all mental entities are physical.

For showing that *Thesis* 3 is very likely true, we must, among other things, take a careful look at the best arguments that have been advanced in favour of the thesis that *all mental entities are physical*. But, actually, there do not seem to be any proposed arguments that are intended to show, in one fell swoop, that *all mental entities* are physical. There are, however, proposed arguments that

are intended to show that all mental *events* are physical (they are arguments for the so-called token-identity theory), or that all mental *properties* are physical (they are arguments for the so-called type-identity theory). I will restrict my attention to the physicalistic arguments referring to *events*, where I take events to include *processes* (events involving change) and *states* (events not involving change, but consisting only in a stable situation).

I claim that there have not been proposed any better *arguments* for the thesis that *all mental events are physical* – in contrast to *propaganda*, by which I mean the mere disparagement of dualism (as "unacceptable," "incoherent," "irrational," "illusory," "motivated by religious prejudice," "anti-scientific," "intellectually unclean," etc.) – than the two arguments that I am going to present below. This is, of course, a rather courageous assertion. Many readers will disagree with it – and it makes it very easy for them to disagree with it. They merely need to cite what they think *is*, or just *might be*, a better argument for the thesis that all mental events are physical (for example: "What about Davidson's argument in 'Mental Events'?") – and, sure enough, I have not even mentioned that argument, let alone shown, in all due detail, that it is not really better than the two to be presented below! I trust that readers such as these will nevertheless find something valuable in this paper, perhaps a reason for being *slightly doubtful* about the truth of physicalism, or a reason for thinking that *Thesis* 3 just *might* be true. Independently of what readers may say, I certainly have to admit that, for all I know, there may have been – or perhaps will be – arguments proposed that are better than those that I am going to present. While I cannot exclude these possibilities (even if I wrote a book of 500 pages instead of a short paper), I adopt the attitude of *wait and see*. Let's wait and see whether anybody has come up, or will come up, with a better argument than those presented below. But there are two requirements: (1) it must be an argument for the conclusion that *all mental events are physical*, or for a conclusion logically stronger than that, and not for a thesis whose sole claim to being physicalism regarding (at least) mental events is that someone or other labelled it that way; (2) it must be an *argument* (with clear premises, a clear conclusion, and a valid logical nexus). For the time being I hold: if the following two arguments for the less general assertion – the assertion that *all mental events are physical* – are not successful, then it does not seem likely that any argument for the completely general assertion – namely, the assertion that *all mental entities are physical* – will be successful. This is a serious situation for physicalists, considering that the dualistic opposition seems to be indeed *better off, both* regarding arguments *and* direct empirical support (concerning this important point, see Sections 4 and 5). But here come the arguments:

The Argument from the Causal Closure of the Physical World

$P_1$1: Every mental event has a physical effect.
$P_1$2: Everything that has a physical effect is physical.
C_1: Every mental event is physical.

This is (the distilled form of) the argument for physicalism regarding events that is most frequently advanced. Great minds do not disdain it (see, for example, Dennett (1991), p. 35, where it is clearly visible). The argument is so popular that more or less corrupted versions of it have found their way into German textbooks on the (Anglo-American) philosophy of mind (see Bieri (1981), p. 5, and Beckermann (1999), pp. 115-117).

The Argument from the Identity of the Functionally Equivalent

$P_2$1: For every mental event there is a physical event that is functionally equivalent with it.
$P_2$2: Functionally equivalent events are identical.
C_2: Every mental event is a physical event.

Though Donald Davidson was not a functionalist, this functionalist argument for physicalism regarding events is basically due to him (see Davidson (1980), p. 179), as will become entirely clear when the concept of functional equivalence is explicated in Section 3 as *causal* equivalence. It has won far less explicit adherents than the Argument from the Causal Closure of the Physical World – a somewhat surprising fact, considering that it is impossible for a physicalist to deny $P_2$1 (as we shall soon see). But the Argument from the Identity of the Functionally Equivalent must be implicitly relied on by any functionalist who also wants to be a materialist (and practically *all* functionalists want to be materialists). In fact, it is closely related to an influential argument "for the Identity Theory" stated long ago by David Lewis (see Lewis (1966), and Lewis (1972)) and, independently, by David Armstrong (see Armstrong (1968)). In Lewis (1994), p. 418, that argument is restated in the following compendious way:

> mental state M = the occupant of the M-role,
> physical state P = the occupant of the M-role (by science),
> therefore M = P.

By unpacking the definite description ("the occupant of the M-role"), this becomes:

> mental state M is an occupant of the M-role,
> physical state P is an occupant of the M-role,

> the M-role has at most one occupant,
> therefore M = P.

Since this is intended to be a general argument, there are two implicit assumptions operative in it: (1) that every mental state, M, shares its functional role – the M-role – with some physical state, P – which (if states and events are counted the same) is just a different way of saying what is asserted by $P_2$1; (2) that at most one state occupies a functional role – for example, the M-role –, or in other words: that states which occupy the same functional role are identical – which (if states and events are counted the same) is just a different way of saying what is asserted by $P_2$2.

2 Critical comment on the Argument from the Causal Closure of the Physical World

As I mentioned, this argument is the one that is most frequently advanced by advocates of physicalism; it is trusted to an extent that would justify for it the name of "rock of physicalism." Its conclusion does, of course, follow logically from its two premises. But neither one of its two premises is such that a denial of it would be irrational, or at least contrary to the "spirit of science."

If by "events" we mean *actual* events: *events that really happen* – and I stipulate that, within this paper, we stick to this meaning –, then $P_1$1 seems highly plausible, but not more than that. It is not inherently irrational to suppose (as those special dualists, the epiphenomenalists, do suppose) that some mental event has no physical effect. But, on the other hand, $P_1$1 has certainly not been refuted by anyone; in particular, no counterexample to it has ever been presented (nor is it likely that there ever will be). I myself sympathize with $P_1$1. But it should be noted that it cannot be justified by the frequently invoked (and time-honoured) principle that only what has an effect exists: *esse est efficere*. Suppose this is true. Nevertheless, the principle leaves the nature of the effect quite unspecified, and hence it supports $P_1$1 to no greater degree than it supports the hypothesis that *every physical event has a mental effect* – and that degree cannot be very high.

$P_1$2 – the hub of the argument – is the *Strong Principle of the Causal Closure of the Physical World*. One should distinguish it carefully from the *Weak Principle of the Causal Closure of the Physical World*:

> Everything physical that is the effect of something is also the effect of something physical.

Obviously, the Strong Principle logically implies the Weak Principle. Equally obviously, the Weak Principle cannot be put in the place of the Strong Principle in the Argument from the Causal Closure of the Physical World. For then the conclusion of this argument will no longer be a logical consequence of its premises; in order to still get the desired conclusion (that all mental events are physical), one must add an extra premise, for example: every effect of something is the effect of *only one thing*, forbidding causal over-determination (cf. Kim (1989)). The addition of *this* premise, although together with the other two premises it does guarantee the desired conclusion, does not help to strengthen the argument: because it states an *additional* assumption which is certainly not of the undeniable kind.[7]

The *Principle of the Causal Closure of the Physical World* is notoriously held by physicalists to be a principle *physics* cannot do without. Sometimes this principle is even held to be entailed by physics. But the first question to ask is this: *which* closure principle do physicalists have in mind? Is it the Strong Principle or is it the Weak Principle? Suppose it is the Weak Principle. Then we must ask: is it true that physics cannot do without the assumption that everything physical that is the effect of something is also the effect of something physical? Would physics – as we know it today, or as it will be at some future point in time – break down if it turned out that something physical is the effect of something, but not the effect of anything physical?

If physics is supposed to stick to the physical (as seems right), then the envisaged violation of the Weak Principle – which is ipso facto also a violation of the Strong Principle – would imply that not all the causes of the physical fall within the purview of physics. But if it were so, would this be the downfall of physics? I suppose not. The only ambition regarding causal knowledge that is inherent in physics, qua science of the physical, is the ambition to find the *physical* causes of the physical – to the extent that the physical has physical causes (and it is not for physics to prejudge that extent in any way). Thus, if some causes of the physical turned out to be non-physical, then the ambition regarding causal knowledge that is inherent in physics would not be frustrated. The only thing that would be frustrated is a *meta*physical bias.

It seems that physics can very well do without the Weak Principle of Causal Closure, and therefore (as a logical consequence) it can also very well do without the Strong Principle. Therefore: the authority of physics cannot be invoked in support of premise $P_1 2$ of the Argument from the Causal Closure of

[7] If motivated by ontological considerations only, then the exclusion of causal over-determination seems arbitrary: there is no good purely ontological reason against causal over-determination. If, on the other hand, the exclusion is motivated by epistemic interests (as it generally is) – mainly, by the interest in obtaining parsimonious explanations –, then one may well ask what epistemic interests can do for ontology. There may be a connection between the two, but it is certainly not obvious or unproblematic.

the Physical World. The assumption of $P_1$2 is not a mandate of physics. Nor is it a mandate of scientific rationality. There is nothing inherently unscientific in the very idea that some causes of the physical are not physical. After all, the non-physical causes of the physical might very well be amenable to scientific inquiry (though not to the inquiry of physics properly speaking). Thus, it appears to be, purely and simply, a matter of metaphysical faith whether one believes in $P_1$2; but if it is a matter of metaphysical faith, then it cannot be deemed irrational to believe otherwise.

It is sometimes held that there is a direct argument from the doctrinal body of physics to $P_1$2. The law of the conservation of energy is invoked as the basis for $P_1$2 by asserting that if something non-physical had a physical effect, then that law would be violated. How so? The idea is this: the non-physical causation of a physical effect would have to lead to an increase in the amount of (physical) energy that is present in the physical world, and such an increase would contradict the law of the conservation of energy.

But, *first*, this law *could* only be contradicted by the non-physical causation of a physical effect if the physical world were a closed system; for what the law says is that *in any closed system* – that is: in any system where there is neither an influx of energy from what is outside the system, nor an outflux of energy into what is outside the system – the total amount of energy neither increases nor decreases. Clearly, the assertion that the physical world is a closed system is not a part of the law of the conservation of energy; nor is it even an assertion of physics qua physics, or of natural science qua natural science; and, certainly, it is not an assertion that is *obviously* true or, if not obviously true, an assertion that has been *shown* to be true. No, the assertion that the physical world is a closed system is, to all appearances, a *metaphysical assertion*. If this is what it is, then believing in this assertion cannot be declared to be a mandate of rationality, let alone of *scientific* rationality.

And, *second*, even if the physical world were a closed system, the law of the conservation of energy would only be contradicted by the non-physical causation of a physical effect if the occurrence of such causation led *indeed* to a change in the total amount of energy that is present in the physical world. But that it would do so is far from obvious. Physics today is ready to countenance physical events that are events of absolute chance, i.e., events without a (sufficient) cause. The existence of such events does not violate the law of the conservation of energy. Nor would this law be violated, I take it, if some of the physical events that are supposed to be events of absolute chance were located in a brain and, in fact, *not* chance events but *caused* by something that is not physical (though still having no physical cause).

But is not $P_1$2 simply an empirically well-confirmed hypothesis? Is it not true that *so far* everything observed to have a physical effect has turned out to be physical? Well, one will hold that this is in fact *not true* if one is *not* a

physicalist. Physicalists and interactionist dualists agree that some mental events have been observed to have physical effects (to the extent that one can observe causation: the having of effects); and the dualists cold-bloodedly add that at least some of these mental events have been observed to be not physical (*not physical* being precisely the character under which those events *appear* to them and to everyone else – at least if one is not already totally subservient to physicalistic prejudice). Thus, for (normal) dualists, $P_1 2$ *is not* empirically well-confirmed but, on the contrary, empirically refuted.

$P_1 2$ is not simply an inductively well-confirmed hypothesis. And we have seen that $P_1 2$ cannot be deduced from the doctrinal body of physics; nor can it be deduced from the doctrinal body of natural science as a whole. Moreover, $P_1 2$, as we have also seen, can neither be justified by the methodological interests of physics nor by the methodological interests of natural science as a whole. $P_1 2$ is just a tenet of a particular metaphysical creed: the creed of physicalism. The question-begging nature of the Argument from the Causal Closure of the Physical World stands now revealed.

3 Critical comment on the Argument from the Identity of the Functionally Equivalent

Prima facie $P_2 1$ might be regarded as being empirically well-confirmed, until one notices, subsequent to conceptual analysis, that it is rather doubtful whether one could ever come up with an instance of empirical confirmation for it – in contrast to $P_1 1$, for which instances of empirical confirmation are easily found *if* one manages to turn the concept of causation into a verifiable empirical concept (which is not easily done; but I will let this pass). $P_2 1$ has the following logical form:

$$\forall x(\Psi(x) \supset \exists y(\Phi(y) \wedge R(y, x))).$$

This makes it seem as if $P_2 1$ were easily confirmable (just like $P_1 1$). But the predicate "y is functionally equivalent with x," which stands in the place of "R(y, x)," is a *defined* predicate, its definition, for events x and y, being the following:

> y is functionally equivalent with x $=_{Def}$ (1) every cause of x is a cause of y, and (2) every cause of y is a cause of x, and (3) every effect of x is an effect of y, and (4) every effect of y is an effect of x.

In other words, events are functionally equivalent if, and only if, they have the same causes and the same effects. The having of the same causes and the same

effects, however, amounts to the truth of four logically unrestricted universal statements (see the above definition). Thus, the presenting of even a single confirmatory instance of $P_2$1 requires the prior verification of four logically unrestricted *all*-statements. And thus, the prospects of confirming $P_2$1 are, from the logical point of view, dim indeed.

But perhaps the set of causes and effects any event has is de facto – though not for logical reasons – *finite*. The implicit assumption that it is finite – perhaps consisting of only two elements: *the (immediate total) cause* and *the (immediate total) effect* of the event – may help to account for the cheerfulness with which most people (even those who are *not* physicalists, even a substance dualist like me: see Meixner (2004)) accept $P_2$1 as something science has shown us to be true, or is at least about to show us to be true. But science is still very far away from showing the truth of $P_2$1, and in fact there is some reason to be pessimistic about it ever doing so.

I suspect, however, that, for physicalists, there is a much more sinister implicit assumption at work here than the one just mentioned: they are positive that for every mental event there is a physical event that is functionally equivalent with it because they implicitly assume that every mental event is *identical* with some physical event (identity entailing functional equivalence, of course). If this is the situation, then the Argument from the Identity of the Functionally Equivalent is rendered *question-begging*. Whatever support $P_2$1 may have, given the conclusion of the above argument of which it is the first premise, $P_2$1 cannot be taken to be supported by the assumption that every mental event is identical with some physical event.

Unlike $P_1$1, $P_2$1 cannot be denied by a physicalist, not even in principle. A physicalist *can* without contradicting herself accept mental events that have no physical effect; she only needs to pay the price for this and accept physical events that have no physical effects (which, of course, *is* a highly implausible assumption, especially considering that the physical events in question would be mental events). But a physicalist *cannot* without contradicting himself accept mental events which are such that no physical event is functionally equivalent with them; for if she accepted this, it would be logically incumbent upon her to also accept that some mental events are not physical, since it can only be true that a mental event has *no* physical event that is functionally equivalent with it (not even itself) if it is itself not physical.

Thus, $P_2$1 is logically entailed by physicalism. This logical fact makes it rather difficult for physicalists not to beg the question in the Argument from the Identity of the Functionally Equivalent. They are bound to have the illusion that $P_2$1 is as good as certain. But in fact they are only seeing $P_2$1 in the favourable light of their *favourite* metaphysical assumption. It is incumbent upon them to adduce support for $P_2$1 which is independent of that assumption. But we have already seen that such support is not easy to come by. This fact

notwithstanding, physicalists are, *qua* physicalists, *condemned* to believe in the truth of $P_2 1$ and to desperately seek confirmations of it. For in whatever degree $P_2 1$ remains unconfirmed, physicalism, too, must remain unconfirmed, which unconfirmedness, if persistent, must be a powerful source of belief-eroding doubt – provided, of course, that one is acting rationally.[8]

Concerning $P_2 2$, it should be noted that it was once suggested by Donald Davidson as a criterion of identity for events.[9] That $P_2 2$ is the Davidsonian identity-criterion for events becomes immediately apparent if we keep in mind the above-presented definition of functional equivalence for events. But we are not concerned here with the question whether $P_2 2$ can indeed serve as a criterion of identity for events (which is doubtful), we are only concerned with the question whether $P_2 2$ can be plausibly regarded *as true*. This is all that matters for the evaluation of the Argument from the Identity of the Functionally Equivalent.

The rationale behind $P_2 2$ is the assumption that events are purely *functional* entities, that all there is to them is to be nodes in a causal net. According to this assumption, an event has no inner being in addition to its causal role; aside from its causal role (which can be profitably identified with the set of its causes and effects), an event is considered to be nothing at all. Thus, $P_2 2$ is a principle of *functionalism*.

Note that $P_2 2$ can be falsified (i.e., rendered false) not only by functionally equivalent events in a causal net which are, nevertheless, *different*, due to their inner nature, but also by *causally isolated* events. Causally isolated events are events that have no causes and no effects. Hence, trivially, all such events are functionally equivalent, and hence, if there happen to be *two* causally isolated events, $P_2 2$ will be false.

But though it can in principle be falsified: *rendered false by counter-instances*, as we have just seen, $P_2 2$ seems to have the great advantage of being not falsifiable *by us*: *not knowable as false by our presenting of counter-instances*. (Karl Popper, of course, would have had his doubts about *this* being an advantage.) In order to falsify $P_2 2$, one has to adduce two *different* events, *a* and *b*, that are nevertheless functionally equivalent. But by the very act of presenting *a* and *b*, veridically and with justification, as *different* events, it seems one is already showing that they are *not* functionally equivalent. For how could a human being, veridically and with justification, present *a* and *b* as different events if they did not have *different effects* on this human being?

---

[8] Here, and also for excluding apparent counterexamples to $P_2 2$ (see below in the main text), the temptation for physicalists "to go eliminativist" is great: if there are *no* mental events, then $P_2 1$ is trivially true – and so is C_2. Some physicalists (for example, the Churchlands and Daniel Dennett) have succumbed to this temptation. But this move – closing one's eyes to reality – can hardly be considered rational.

[9] See Davidson (1980), p. 179. Significantly, it is pointed out by Davidson that his criterion is useful for establishing the identity between a pain and a complex physiological event.

But this rhetorical question, demanding the answer "yes," presents something as being obvious which just isn't obvious at all. I can, veridically and with justification, present 1 and 2 as different numbers, although they do not have different effects on me (since they do not have any effects on me at all). I can distinguish them nonetheless, by cognizing – in an intersubjectively accessible way – their inner nature. I can do this because not all cognition rides on the back of causation which originates in what is being cognized. What is true of the numbers 1 and 2 may also hold true of one of my mental events, *a*, and a physical event, *b*, which is a functional equivalent of *a*: though *a* and *b* have the same causes and the same effects, I may well be able to present them as different, doing so veridically and with justification.

Here is *how* (and I am merely repeating a story that has, one way or other, been often told). Suppose science has advanced to the point that *a* and *b* are shown to be functional equivalents of each other. (So far, this has not been shown for any mental event *x* and physical event *y*.) But the mental event *a* – a certain experience of mine, say, my seeing and touching of a red wooden stick, put in perfectly clear water, appearing at once bent to my vision and straight to my touch – is immediately present to me: in a non-causal way, since I am its intrinsic subject, with certain subjective traits and with some illusory intentional content. The event *b* – a neurophysiological occurrence, presumably – is not present to me in that way at all: there is nothing intrinsically subjective about it.[10] Moreover, the illusory intentional content of *a* – taken at face value: in all its richness and vividness – makes it quite impossible to locate *a* within physical reality, be it inside or outside brains. Though functionally (i.e., causally) equivalent with *a*, *b* simply cannot fill the place of *a* (entirely). This is the way things *seem* to me (and also to physicalists, I presume, at least *prima facie*). Thus, trusting appearances, I am entirely justified in passing the judgment that *a* and *b* are, though functional equivalents of each other, *different* events. Physicalists, in turn, not trusting appearances (but they certainly must trust *some* appearances, or there could not be any knowledge for them), are justified in pointing out – as they do, as a rule –that I *might* be mistaken (i.e., that one is rationally allowed to take the degree of credence for my being mistaken to be greater than 0); after all people often thought that X and Y were different, though in fact they were identical. (There are a lot of stock examples for that, beloved by physicalists, usually taken from the resplendent history of scientific progress.) Indeed, I might be mistaken in thinking that *a* and *b* are different; but this alone does not take away my being

[10] Do I merely think so because I am already a dualist and see the world in the light of my metaphysical bias, not realizing that I do so? While I might certainly be moved by motives hidden to me, merely uttering a suspicion in this regard does not justify that suspicion – a point not generally appreciated by physicalists.

justified in asserting that they are different. For taking away my justification in this regard, physicalists must *show* that I am *in fact wrong* about *a* and *b* being different (and this is not already done by showing that I *might* be wrong[11]). Since appearances are on my side, the burden of proof lies with them. But how could physicalists *show* that *a* and *b* are not only functional equivalents of each other but also identical (without begging the question regarding physicalism)? Only by assuming a general principle that sanctions the relevant inference of identity from functional equivalence, or in other words: by maintaining $P_2$2 in the teeth of an apparent counterexample to it. But they certainly do not have any more rational right to hold on to $P_2$2 than I have to hold on to the apparent counterexample to $P_2$2. In view of that counterexample, *Ockham's razor* – "it's parsimonious to identify functionally equivalent events" – or *inference to the best explanation* – "the best explanation of the functional equivalence of events is that they are, in fact, identical" – are problematic procedures and cannot be safely relied on. Thus, jumping from functional equivalence to identity remains a *blind* leap of faith – and in the case of the mental event *a* and the physical event *b* a foolhardy one, I should say.[12]

Summing up, my objection to the Argument from the Identity of the Functionally Equivalent is this. Either $P_2$1 is wrong and for some mental event there is no physical event that is functionally equivalent with it – then physicalism stands refuted –, or, indeed (though we are still very far away from having shown this), $P_2$1 is true and for every mental event there is a physical event that is functionally equivalent with it. But then my mental event *a* and the physical event *b* that is its functional equivalent will still appear to be different events, and therefore $P_2$2 will seem to be "counterexampled." $P_2$2 is not an analytic principle; in fact, it has no support which is not of a question-begging nature: it has no support which is not obviously guided by materialist metaphysical bias. Therefore: if $P_2$2, in the envisaged situation, is maintained in the teeth of the evidence against it, then this move is not more, but certainly less reasonable than maintaining, in the envisaged situation, what that evidence straightforwardly tells us: that $P_2$2 is false.

4 Is the opposition no better off?

[11] Unfortunately, it is not entirely unnecessary to make this parenthetical remark.

[12] Kim writes: "[A] certain instability exists in a situation in which two distinct events are claimed to be nomologically *equivalent* causes or explanations of the same phenomenon; stability is restored when equivalence is replaced by identity or some asymmetric relation of dependence." (Kim (1989), p. 246.) I fail to discern the "certain instability" Kim is talking about. And even if it existed, it would be an instability in *our theorizing*. Can we draw ontological conclusions from that? I should think, we cannot.

I have argued that physicalists, if they offer arguments for their position regarding mental events at all, do not have better arguments on offer than the two arguments discussed. But perhaps the dualistic opposition is no better off? Perhaps the arguments of dualists *against* physicalism regarding mental events are just as bad as the arguments of physicalists *for* physicalism regarding mental events?

If one does not want to install dualism instead of physicalism on the throne of metaphysics, this objection is strictly irrelevant. But let me make it relevant: *Descartes*, for one, would have liked to see dualism instead of physicalism on the throne of metaphysics. In fact, he thought dualism is the only position that has any right to be there. So what did he have to offer in support of his claim? Here is a variant of the Cartesian argument, trimmed to its bare essentials:

$P_3$1: There is a mental event for which it is possible that it exist without anything physical existing.
$P_3$2: There is no physical event for which it is possible that it exist without anything physical existing.
C_3: Some mental event is not physical.

Most people believe that the notion of possibility that is operative in this argument is crucial for assessing it. Pick, therefore, the sense of possibility according to which almost everybody agrees that $P_3$1 is true: interpret "possible" to mean as much as "epistemico-logically possible," that is: *possible in the broadest sense, given the conceptual framework that fits our current knowledg*e. Employing this sense of "possible," $P_3$1 seems incontrovertibly true, and not only to dualists (in sharp contrast to the situation obtaining when one interprets "possible" to mean as much as "metaphysically possible"). But in interpreting $P_3$2 the same interpretation of "possible" must be used as in interpreting $P_3$1 (or else one commits the fallacy of equivocation). The intriguing thing is that $P_3$2, with the epistemico-logical interpretation of "possible" in place, certainly does not seem to be less true than $P_3$1, with that same interpretation of "possible" in place.

This looks like a desperate situation for physicalists. For the conclusion that is unacceptable for them – if it is not unacceptable for them, how can they still call themselves "physicalists"? – follows logically from the premises, since the logical structure of the argument is an uncontroversially valid inference pattern:

$P_3$1 $\exists x[\Psi(x) \wedge K(x) \wedge \Diamond(E(x) \wedge \neg\exists y(\Phi(y) \wedge E(y)))]$
$P_3$2 $\neg\exists x[\Phi(x) \wedge K(x) \wedge \Diamond(E(x) \wedge \neg\exists y(\Phi(y) \wedge E(y)))]$

C_3 $\exists x[\Psi(x) \wedge K(x) \wedge \neg\Phi(x)]$

What is to be done now? For physicalists, if they want to hold on to the negation of the conclusion, there is no choice but to deny the premises. Perhaps they will try to avoid this move by tacitly redefining themselves as physicalists in such a way that, qua physicalists, they are *not* committed to believing that all mental events are physical events. But I have argued that redefining physicalism in this way amounts to *giving up* physicalism.

We must, therefore, ask: *how* can physicalists deny the premises $P_3 1$ and $P_3 2$ (*as interpreted*)? Suppose physicalists single out $P_3 1$ for being denied. But is it at all plausible that, for *every* mental event, it is *epistemico-logically impossible* that it exist without anything physical existing? It is widely accepted that conceivability is sufficient for *epistemico-logical* possibility. I (and I am certainly not alone in this) can conceive of my present mental state existing without anything physical existing; hence it is *epistemico-logically possible* that it exist without anything physical existing.

The prospects of denying $P_3 2$ appear to be even dimmer. The idea of a physical event for which it is epistemico-logically possible that it exist without anything physical existing seems preposterous – just as preposterous as the idea of a human being (a number, etc.) for which it is epistemico-logically possible that it exist without any human being (number, etc.) existing. However, it is important to realize that *not all* statements of the form $\neg\exists x[\Phi(x) \wedge K(x) \wedge \Diamond(E(x) \wedge \neg\exists y(\Phi(y) \wedge E(y)))]$ are true: "There is a first man on the moon who is American and for whom it is epistemico-logically possible that he exist without any first man on the moon existing" is certainly a true statement, and therefore its negation – which is a statement of the form $\neg\exists x[\Phi(x) \wedge K(x) \wedge \Diamond(E(x) \wedge \neg\exists y(\Phi(y) \wedge E(y)))]$ – is *false*. Nevertheless, a statement of the form $\neg\exists x[\Phi(x) \wedge K(x) \wedge \Diamond(E(x) \wedge \neg\exists y(\Phi(y) \wedge E(y)))]$ will turn out true – provably so, no matter how the possibility operator $\Diamond$ is interpreted, as long as the following utterly elementary inference-pattern of modal logic stays applicable: $\Diamond A, \neg\Diamond\neg B \rightarrow \Diamond(A \wedge B)$ – for every predicate $\Phi(x)$ for which one has: $\forall x(\Phi(x) \supset \neg\Diamond\neg\Phi(x))$. Whatever relevant interpretation of $\Diamond$ is chosen, predicates $\Phi(x)$ for which $\forall x(\Phi(x) \supset \neg\Diamond\neg\Phi(x))$ is true are, for example, "x is a human being," "x is a number," *and* – it seems undeniable – "x is physical."

Is it in fact undeniable that for everything physical it is not epistemico-logically possible that it be not physical? This table is physical, and it is certainly not, in any sense, possible that it be not physical. This brain is physical, and it is not, in any sense, possible that it be not physical. This hydrogen atom is physical, and it is certainly not, in any sense, possible that it be not physical. And so on. Whatever room is left for denying $\forall x(\Phi(x) \supset \neg\Diamond\neg\Phi(x))$, with $\Phi(x)$ being "x is physical" and $\Diamond$ designating epistemico-logical possibility (it doesn't seem much room to me), it is all the room

physicalists have for escaping the Cartesian argument without appearing utterly unreasonable about the matter.

It is important to fully appreciate this. Otherwise $P_3 2$ can seem very easily deniable – at least for physicalists. For take some mental event, say, an experience of pain. Many physicalists will blithely acknowledge that it is *epistemico-logically* possible that this pain-experience exist without anything physical existing (thereby accepting $P_3 1$). But, of course, they will also hold that the pain-experience is *physical*. Hence, they conclude that for some *physical* event – the pain-experience – it is epistemico-logically possible that it exist without anything physical existing – *contradicting* $P_3 2$.

Well, nothing easier than to use the negation of the conclusion of a logically valid argument to "refute" one of its premises! In fact, the use of this dialectical procedure would be impeccable in the present case if the refuted premise had no other support than that it serves to establish the desired conclusion; if this were so, one would have to judge the argument *question-begging*. However, as we have just seen, $P_3 2$ does have support that is independent of the conclusion it serves to establish: it is grounded in modal logic and the highly plausible principle that *for everything that is physical it is epistemico-logically impossible that it be not physical*.

5 Arguments don't matter?

The *sophisticated* physicalist philosopher – I need not provide any names, for the *sophisticated* physicalist philosopher is very common, and becoming more common every day – will be utterly unimpressed by all of this: both by my criticism of the best arguments for event-physicalism, and by my presentation of a strong argument for the negation of event-physicalism. For the *sophisticated* physicalist believes that arguments don't matter; what matters are *empirical facts*. Moreover, the sophisticated physicalist believes that the empirical facts are *all* in favour of physicalism, and that there are *none* in favour of dualism.

However, one can make a case for the position that physicalism is *refuted* by empirical fact. As follows. There are *illusory experiences*: experiences real enough as experiences, but with illusory contents (an example has been given in Section 3). The existence of illusory experiences is an empirical fact, and it seems that they just cannot be fitted into a purely physical world, neither from the start: without prior reduction, nor after any conceivable reductive process. Illusory experiences seem to be *irreducibly non-physical*.[13]

[13] For details, see Chapter 4 of Meixner (2004). See also the considerations surrounding the example given in Section 3.

Dennett, for one, is much exercised by this difficulty for physicalism and is led to an – implausible – eliminativist position regarding what he calls "real seemings."[14]

Moving on to an objection against sophisticated physicalism that specifically addresses its rejection of philosophical argument, one may well ask: how could empirical facts all by themselves – without argument, and hence without theory – demonstrate anything? Is it that empirical facts, all by themselves, somehow create a certain atmosphere in which only one *Weltanschauung* is possible? But this question, I fear, is lost on the sophisticated physicalists. Perhaps the following considerations will be more effective (although I doubt it). What is revealed by empirical research – in ever greater detail and accuracy (but there is no saying whether we will ever reach *completeness* regarding detail and accuracy; there is some indication that one has been far too optimistic in this regard) – is the manner in which neurophysiological conditions are necessary, respectively sufficient, conditions for the occurrence of mental events. But in no way does empirical research imply that mental events are nothing else than the neurophysiological events that condition them. On the contrary, that mental events are, at least provisionally, *something else* than physical events is a methodological premise of brain-consciousness research.[15] Suppose, then, that brain-consciousness research develops in the best possible way. Would that mean that one reaches, purely as a result of empirical research, the insight that all mental events are certain physical events? No, all one would reach, purely as the result of empirical research, is this: for each mental event, one is able to provide the necessary and sufficient physical condition. This is the best possible outcome of empirical brain-consciousness research. Everything that goes beyond this is *metaphysics*. I have no objection to metaphysics. On the contrary, I believe that it is an absolutely indispensable part of philosophy. But what might be expected of philosophers is to be able to distinguish when they are doing metaphysics, and when not, and when they are doing metaphysics in a philosophically responsible way, and when not.

In summing up, I repeat my appeal: it is time to return to intellectual honesty in the philosophy of mind. The honest position is to admit that dualism is *rationally* at least on a par with physicalism. This position is compatible with being an *agnostic* regarding these basic metaphysical positions. It is *also* compatible with being a *believer* in physicalism, respectively dualism. What the materialist believers, as a rule, have not realized (or refuse to realize) is that they have a certain metaphysical *faith*: a belief that they accept on other than rational or scientific grounds – which situation, it should be noted, *does not* by itself entail that they are not rationally permitted to have this belief. But,

[14] See Dennett (1991).
[15] On methodological dualism, see Chapter 7 of Meixner (2004).

speaking of scientism instead of materialism (*de facto*, the difference is immaterial), Bas van Fraassen has put very well what would be *rationally right* also for the materialist *believer*:

> Commitment to the scientific enterprise does not require [scientism]. If anyone adopts such a belief, he or she does it as a leap of faith. To make such a leap does not make us *ipso facto* irrational; but we should be able to live in the light of day, where our decisions are acknowledged and avowed as our own, and not disguised as the compulsion of reason. (Van Fraassen (1991), p. 17.)

Uwe Meixner
Institute of Philosophy, University of the Saarland,
P.O. Box 151150, D-66041 Saarbrücken, Germany
Email: u.meixner@mx.uni-saarland.de

References

Armstrong, David (1968), *A Materialist Theory of Mind* (London: Routledge).
Bieri, Peter (ed.) (1981), *Analytische Philosophie des Geistes* (Königstein/Ts.: Hain).
Beckermann, Ansgar (1999), *Analytische Einführung in die Philosophie des Geistes* (Berlin/New York: De Gruyter).
Davidson, Donald (1980), 'The Individuation of Events', *Essays on Action and Events* (Oxford: Clarendon Press), pp. 163-180.
Dennett, Daniel C. (1991), *Consciousness Explained* (Boston/New York/London: Little, Brown and Company).
Kim, Jaegwon (1989), 'Mechanism, Purpose, and Explanatory Exclusion', *Philosophical Perspectives* 3, *Philosophy of Mind and Action Theory*, edited by James E. Tomberlin (Atascadero, CA: Ridgeview), pp. 77-108.
Kim, Jaegwon (1993), *Supervenience and Mind* (Cambridge: Cambridge University Press).
Kim, Jaegwon (2001), *Mind in a Physical World*, fifth printing (Cambridge, Mass.: The MIT Press).
Lewis, David (1966), 'An Argument for the Identity Theory', *The Journal of Philosophy* 63, pp. 17-25.
Lewis, David (1972), 'Psychophysical and Theoretical Identifications', *Australasian Journal of Philosophy* 50, pp. 249-258.
Lewis, David (1994), 'Reduction of Mind', *A Companion to the Philosophy of Mind*, edited by Samuel Guttenplan (Oxford: Blackwell), pp. 412-431.
Meixner, Uwe (2004), *The Two Sides of Being* (Paderborn: mentis).
van Fraassen, Bas (1991), *Quantum Mechanics. An Empiricist View* (Oxford: Clarendon Press).