Uwe Meixner

New Perspectives for a Dualistic Conception of Mental Causation

1. Nonphysical Causation of Physical Events?

It is a curious fact of the history of philosophy that so many philosophers have complained about the incomprehensibility of nonphysical mental causation of physical events, considering that most of the many philosophical conceptions of causation on offer (i.e., regularity theories, counterfactual theories, and probabilistic theories) do not give any grounds for supposing that there is anything particularly incomprehensible about the nonphysical causation of physical events. It should be noted that the principles of causal closure of the physical world — constantly invoked against the nonphysical causation of the physical — are neither principles of the logic of causation nor principles of physics, but postulates of materialist metaphysics. As such, the closure principles are begging the very question which is at issue.

But does not physics itself tell against the nonphysical causation of the physical? It does not. In the first place, it is rather unclear whether the concept of causation is indispensable for physics. The concept of *force* is the place where causation must come into physics if it comes into physics at all, and it must be admitted that there is much causal talk surrounding that concept, even among physicists. But this causal talk seems to be entirely due to extra-scientific motivations and associations. For, regarded purely as a concept of *physics*, the (net) *force* a particle is subjected to at a time t is definable (though not usually *defined*) as the (net) change of momentum that the particle undergoes

Correspondence:
Uwe Meixner, Institute of Philosophy, University of Regensburg, D-93040 Regensburg, Germany. *Email: uwe.meixner@psk.uni-regensburg.de*

Journal of Consciousness Studies, **15**, No. 1, 2008, pp. 17–38

in an infinitesimally small interval of time surrounding t,[1] in analogy to the definition of the particle's (net) acceleration at t, which is defined as the (net) change of velocity that it undergoes in an infinitesimally small time-interval surrounding t. There certainly seems to be not a glimpse of causation in the suggested definition of *force*.

It is alleged again and again that the nonphysical causation of physical events is bound to violate received physics because it, allegedly, entails the violation of the law of the preservation of energy, or the violation of the law of the preservation of momentum. Repetition does not make false allegations any less false. First, in physics, the mentioned preservation laws are always asserted under the condition that the physical system with regard to which they are asserted is a so-called *closed system*: that no energy or momentum is coming into the system from entities that are outside of it, or is going out of the system to entities outside of it. Now, physics is silent on the question whether *the entire physical world* is a closed system. Moreover, it does not seem to be an analytic truth that the physical world is such a system. It follows that in order to have the nonphysical causation of physical events conflict with the preservation laws, it is necessary to go beyond physics and to assume the *metaphysical* hypothesis that the physical world is a closed system.

Making this assumption is a necessary condition for obtaining a conflict; but, note, it is not a sufficient one. For suppose, for the sake of the argument, the physical world were indeed a *closed system*. Consider then: does the occurrence of an instance of nonphysical causation of a physical event necessarily entail that the sum total of energy or of momentum in the physical world is any greater or smaller than before — *in spite* of the physical world being a closed system, as we have supposed for the sake of the argument? Suppose the instance of nonphysical causation we are considering is due to a subjective experience, which, in my view,[2] is by natural (or nomological) necessity *causally equivalent* with a brain state, meaning that it has the same causes and the same effects as that brain state. It is evident that *this* kind of nonphysical causation, which is entirely in step with physical causation, need violate neither the law of the preservation of energy

[1] This definition follows Newton's original formulation of his Second Law of Motion (see Newton, 1962, p. 13; what we call 'momentum', Newton, 1962, p. 1, calls 'the quantity of motion'). Since it does not presuppose the constancy of mass, it is more general than the usual definition of force, according to which force is the product of mass and acceleration (at a given moment of time). It is, however, equivalent to this latter definition if constancy of mass is presupposed, as was done by Newton (but not by Einstein).

[2] For details, see Meixner (2004).

nor the law of the preservation of momentum *if* it is true that physical causation violates neither law. And everybody agrees that physical causation violates neither of these two laws (the physical world being a closed system notwithstanding). Thus, there is no objection that comes from the direction of physics against the nonphysical causation of a physical event if this nonphysical causation is taken to be due to the — nomologically established — causal equivalence between brain states and nonphysical subjective experiences.

Objector: But if nonphysical subjective experiences have causal consequences and are causally equivalent to brain states, then this will imply causal over-determination. This seems to me a serious drawback.

Response: You should not let yourself be misled by the word, 'over-determination'. Like the word 'over-reaction', it suggests that something untoward is going on. I would prefer the neutral expression 'nomologically coordinated causation' for designating the causal situation that due to the laws of nature an event A that is a cause of event C cannot be a cause of C without an event B that is distinct from A being also a cause of C, and vice versa (regarding A and B). The manifest image of the psycho-physical relationship suggests that nomologically coordinated causation in fact occurs. It certainly cannot be ruled out on *a priori* grounds.

2. Purely Nonphysical Causation of Physical Events *or:* Free Nonphysical Agency

But what about the nonphysical causation of physical events *without* equivalent physical causation, say, without *any* accompanying physical causation at all? Would not the occurrence of nonphysical causation of physical events without accompanying physical causation get into conflict with physics? It would not, not even under the metaphysical supposition that the physical world is a closed system: because an instance of nonphysical causation of a physical event without accompanying physical causation would leave the sum total of energy and momentum unchanged. It would merely involve a redistribution of energy and momentum. Redistributions of energy and momentum are, of course, happening constantly, and normally, it seems, one need not invoke nonphysical causation for having them come about. But, as most modern physicists hold, at least some of these redistributions are *not* determined by the energy/momentum distributions of the past.[3] If

[3] Most physicists are following the lead of Max Born who wrote in 1926 that the paths of particles 'are determined only insofar as they are constrained by the principle of energy and momentum conservation; apart from this, the value distribution of the ψ-function

this is true, then the physical past leaves a lacuna of determination that need not be left entirely to chance, but that can be, at least partly, filled by *additional* determination coming purely from a nonphysical source. In an indeterministic physical world, there *is* room for the nonphysical — specifically, the nonphysical *mental* — causation of physical events without accompanying physical causation.

Let me call such causation *purely nonphysical* mental causation of the physical. Purely nonphysical mental causation of the physical, if it occurred, would not originate in subjective experiences, since subjective experiences are causes only in unison with their physical causal equivalents (at least in my eyes). Purely nonphysical mental causation of the physical would originate in the *mental subject*, in the nonphysical individual, wholly present at each moment of its existence, which is the centre of consciousness: in the nonphysical substantial *self*.[4] Since purely nonphysical mental causation of the physical presupposes physical *indeterminism* and originates in a substantial *nonphysical mental agent*, I will also call this kind of causation *free nonphysical agency*.

Supposing that free nonphysical agency exists — physicalists, of course, do everything to make this supposition appear absurd — the fundamental question to be answered by dualists is this: how does free agency fit naturally into a dualistic framework — in such a manner that it does not get into conflict with anything we know about the brain, perhaps even in such a manner that free agency is positively supported by some things we know about the brain?

As far as we bodily existing human beings *know*, the nonphysical mental subject does not exist without a functioning brain.[5] But although the nonphysical mental subject, as far as we know, depends for its nonphysical existence on the brain, it is not a superfluous ghostly excrescence of the brain, but has certain *evolved functions*, via the brain, for the organism as a whole. Its first function is that of being the centre of another nonphysical product of the brain: of consciousness. A mental subject is a subject of consciousness. Its second function is that of being a free agent, acting, via the brain and body, on the

---

determines only the probability that a particle will follow a particular path' (cited and translated in Torretti, 1999, p. 333).

[4] In Meixner (2004), many reasons (in my view, *good* reasons) are provided why *the mental subject*, *the self*, should be conceived of as a nonphysical substance. I cannot repeat these reasons here, but they have to do with the nature of consciousness, without reference to the nature of action.

[5] It is metaphysically possible that the mental subject exist without a functioning brain, but in the normal course of nature we know of this is never actually the case.

behalf of the organism, in the service of its survival (or at least its well-being), *within a certain scope of prior indetermination*.[6] In addition to being a subject of consciousness, a mental subject is a subject of free agency — of free agency that is *guided* by the (non-determinative) information provided to it in consciousness.

In the light of this last remark, it emerges that the first function of the mental subject — its being the subject of consciousness — is subordinate to its second function — its being the subject of free agency. And incidentally, we can now see what consciousness is *good for* (taking this seemingly teleological question, as evolution theorists are wont to do, in the following *non-teleological* sense: what is the reason for the persistence of consciousness in the course of evolution): it effectively provides the subject of consciousness with the (non-determinative) information needed for acting freely and successfully on the behalf of the organism, the organism to which consciousness (precisely speaking: a *particular* consciousness) and its subject (a *particular* subject of consciousness) are connected; cf. Meixner (2004; 2006).[7] As Michael Polanyi put it several decades ago:

> Descending therefore from the person of a great man down to the level of the newborn infant and beyond that to the lowest animals, we find a continuous series of centres whose a-critical decisions account

[6] *Avoidance* is one important type of antecedently undetermined physical action that a mental subject freely undertakes for its organism. It will hardly be considered surprising that this view of avoidance collides with the view presented in Dennett (2003, p. 60): 'If we want to make sense of the biological world, we need a concept of avoidance that applies liberally to events in the history of life on Earth, whether or not that history is determined. This, I submit, is the *proper* concept of avoidance, as real as avoidance could ever be.' I agree with Dennett that we need a concept of avoidance that applies liberally to events in the history of life on Earth. But if we want to make sense of the biological world, we precisely *cannot* assume that that history is determined. I submit, we cannot make sense of *conscious* life — an important part of the biological world — without assuming that the history of life on Earth is *not* determined (for more on this, see Meixner, 2006). We therefore need an *incompatibilist* concept of avoidance, which, incidentally, is also the normal concept: one cannot normally say that a *determined* object *avoids* anything.

[7] In contrast to what is argued in Meixner (2006) and Meixner (2004), Chalmers (1996, p. 120) believes that consciousness does not serve a physical function that could not be achieved without it; consequently, he is sceptical about an evolutionary explanation of consciousness. But Chalmers erroneously believes that *the logical possibility* that consciousness achieves in the physical world merely what can be achieved without it (in other words, *the logical possibility* of a zombie twin) shows the evolutionary irrelevance of consciousness. Not so; something more than a mere logical possibility would be required for that. Dretske (1995, p. 122), on the other hand, holds that 'the function of conscious states is to make creatures conscious — of whatever they need to be conscious to survive and flourish'. For him, the non-redundantly advantageous function of consciousness in the evolutionary process is obvious. His physicalistic stance, however, blinds him to the real force of the problem which he formulates clearly enough a few pages earlier (ibid., p. 119): 'What use is experience in cognition if the same job (the processing of information needed for the determination of appropriate action) can be achieved without it?'

> ultimately for every action of sentient individuality. Thus the personal pole of commitment retains its autonomy everywhere, exercising its calling within a material milieu which conditions but never fully determines its actions (Polanyi, 1969, p. 397).

3. The Brain as an Apparent DOMINDAR

The difficult question is how the nonphysical mental subject manages to do all this. If there is an answer, it must be provided by the brain. I maintain that the brain is, among other things, (1) an instrument for the detection of macroscopic indetermination in the environment of the organism (which environment includes, as its limit, the organism itself) and (2) an instrument for restricting the detected macroscopic indetermination to the advantage of the organism. In short, I maintain that the brain is a DOMINDAR:

Detector Of Macroscopic INDetermination, And Restrictor.

This is a *bold* assertion because it has not seemed to most philosophically tuned people that there is enough macroscopic indetermination in the physical world[8] to be detected or restricted by anything. This, I believe, is a false impression.

But let me first show that the brain certainly *seems* to be a DOMINDAR. Suppose someone, George (precisely speaking: George-in-the-body), fleeing from his deadly enemies, comes to a crossroads. What is his brain doing? It prominently presents — in the foreground of consciousness, to the subject of George's consciousness, which in fact is George himself, I maintain — four alternative items as things he could do within the immediate future (and in the background of consciousness, an indefinite number of further alternative items as things he could also do): turn back, turn right, turn left, go straight ahead. If this presentation of alternatives of action is veridical — and it certainly seems veridical to George (and would seem veridical to us if we were in George's place) — then George's brain has served as a detector of macroscopic indetermination in the environment of its organism; for the presentation in question is veridical only if at the time of the presentation, given the entire physical past and all the laws of nature, the organism can indeed move in one or another of four alternative ways (at least!): turn back, turn right, turn left, go straight ahead.

[8] Indetermination is of course not *in* the physical world in the sense of being more or less literally a part of it. The quantity of indetermination *in* the physical world at a time *t* depends on the quantity of physical and physically possible further world-courses that veer away from each other after time *t*.

Objector: I reject this interpretation of what George is presented with in experience. George is not presented with four alternative actions he can realize, but with four alternative actions each of will be realized depending on his choice. The latter interpretation obviously allows George's experience to be veridical even if determinism is true, since George's choice may itself be perfectly determined.

Response: The second interpretation of what George is presented with in experience collapses into the first interpretation if one adds to the second interpretation: George experiences that he has four choices open to him. I claim that George indeed experiences (i.e., is under the impression) that he has four choices open to him. In fact, this experience is contained in his experience of having four alternative actions open to him. On pain of infinite regress, it is not possible to argue once again: George is not presented with four alternative choices he can make, but with four alternative choices each of which will be realized depending on his choice.

Objector: This amounts to claiming that George, before acting and choosing, experiences his alternative possible actions and choices of action to be causally undetermined.

Response: Precisely. And this experience is veridical only if at the time of the experience, given the entire physical past and all the laws of nature, George's organism can indeed move in more than one way. Thus the experience betokens (veridically or not) macroscopic *physical indetermination*. It also betokens *nonphysical psychological indetermination*, since a choice of action is causally undetermined at a time only if it is still open given *all* antecedent factors (*including* the inner — *psychological* — states of the person, whether physical or *nonphysical*). Nonphysical indetermination, however, is not my concern here.

And George's brain not only appears to be a detector of macroscopic indetermination in the described situation, it also appears to be the restrictor of this indetermination. For once George has decided what to do (on the basis of the conscious experiences his brain makes him have), his brain will implement his decision and accordingly appear to *restrict* the previously apparent indetermination in the environment of George's body. Say, George effectively decides to go straight ahead, likely enough in the light of a brief rational deliberation, taking into account, say, the high probability of a helicopter waiting for him one mile ahead; then it is his brain that makes George's body (and with it George-in-the-body) *go straight ahead* (in the way so well explored by neurophysiology), excluding (or 'closing') thereby all the other alternatives for George and his body that previously appeared to be open to him at this particular juncture of his career.

Thus, it must be concluded that the brain *seems* to be a DOMINDAR. But is it in fact? If the macroscopic physical world is a deterministic world, or practically deterministic world, then the appearance that the brain is a DOMINDAR is an illusion;[9] for then there is certainly not enough indetermination in the macroscopic physical world to be either detected or restricted. But, if it is assumed that macroscopic determinism rules in the physical world, one may well ask why the brain in each moment of conscious existence presents to the subject of consciousness alternatives of action which that subject does not in fact have, systematically misleading it. *If* macroscopic determinism rules in the physical world and we nevertheless for some reason have to have consciousness, why then do we not at least have a consciousness that truthfully tells us in each moment of conscious existence: *this*, and *this alone*, is what I *must* do? I have not seen a plausible answer to this question.

> *Objector*: But in fact there is an easy answer to it: the decision of the person is part of what determines the course of events. The person cannot know his or her own decision in advance before deliberation. This epistemic point is perfectly compatible with the assumption that the result of the decision and the decision itself is nonetheless causally determined by antecedent factors.
>
> *Response*: We have agreed above that George experiences his alternative possible actions and choices of action to be causally undetermined, that is: as being not determined by *all* antecedent factors, where the phrase 'by all antecedent factors' must, in reason, mean: *by all factors antecedent to his decision*, since he also experiences — i.e., is under the impression — that which choice of action he actually makes and which course of action he actually takes will (though previously undetermined) be determined by his decision, whatever that decision will be. This, in total, is what George experiences, call it: *the experience of liberty*. And if determinism is assumed to be true, then this experience of George and the similar experiences we all have in practically all the moments of our conscious lives must be regarded as illusory and need an explanation of why they nevertheless occur to this massive extent. Have you offered such an explanation? You have sketched a psychological mechanism that with some plausibility produces the experience of liberty even if determinism is true — *only* with *some* plausibility, since not always when we do not know how things will turn out the impression arises in us that it is undetermined how they will turn out. It must also be pointed out that one can have — and often has in fact — the experience of liberty even though one *knows* with perfect certainty in advance which choice of action one is going to make and which course

[9] The brain may still be a *potential* DOMINDAR — but a potential DOMINDAR that is never actualized.

of action one is going to take (say, by having made up one's mind regarding the eventuality in question long beforehand). Your 'epistemic point', therefore, falls short of explaining the experience of liberty. Moreover, my demand for explanation did, as a matter of fact, not require an answer to the question of *how the experience of liberty is possible even under determinism* (this is the question you addressed), but it required an answer to the question of *what is the point (mainly, the biological point) of the experience of liberty even under determinism*. I still haven't seen a plausible answer to *that* question.

Objector: How can one already know — I mean, *know* — how one will decide and still be at liberty — still be undetermined by all factors antecedent to one's decision — regarding how one will decide?

Response: Now you are changing the subject.

4. The Brain as an *Instrumental* DOMINDAR and the Libet-Experiment

This is the appropriate place for briefly addressing what the much-discussed Libet-experiment of the 1980s means for the brain's being a DOMINDAR. A detector of macroscopic indetermination and restrictor can be such a thing in two ways: *in its own right*, or *instrumentally for something else*. As I have presented matters, the brain is — among many, many other things, of course — an *instrumental* DOMINDAR *for something else*, namely, for the nonphysical self, which is at once the subject of consciousness and of agency. In my opinion, the brain is an instrument of detection and restriction of indetermination for that self, and not in its own right. This view of the matter has the advantage of *not* turning consciousness and self into phenomena that are superfluous from the biological point of view.

But it does have the disadvantage that it is vulnerable to a standard interpretation of the result of the Libet-experiment. This experiment is standardly taken to show that the brain does not wait for the self to initiate action, that it initiates action on its own, the self merely echoing the brain's decision. If this were the correct view of the matter, then the brain *could* still be a DOMINDAR — nothing in the standard interpretation of the result of the Libet-experiment tells against that. But it would have to be a DOMINDAR *in its own right*, and *not* instrumentally for the self.

Fortunately, the standard interpretation of the result of the Libet-experiment is by no means forced upon us. The standard interpretation is based on the problematic assumption that the moment when the self decides what to do is identical with the moment it becomes conscious of deciding what to do. Suppose the self's decision is in fact, as it

should be, infinitesimally prior to the instrumental initiation of action by the brain, but the self becomes *conscious* of deciding only after the brain has already instrumentally initiated action, and therefore only after the self has already decided. Is this an absurd supposition? It is not. For making an informed decision, the self needs to be conscious of the facts relevant to the decision prior to making the decision; but for making the decision, and for making it in an informed way, the self certainly does not need to be conscious of making the decision at the very same time it makes it.

Being conscious of (presently) making a decision is not relevant to making the decision, neither regarding the *intending* of the decision (of course not), nor regarding its being actually made. The consciousness of making a decision has a different role to play in the economy of action, a role for which it is not necessary that the consciousness of making a decision occur at the very time when the decision is being made. Well, what *is* that role? It is this: the fact that we have decided so-and-so is likely to be in its turn something we need to be informed of in order to make further informed decisions; but for remembering that we have decided so-and-so, we must have been conscious of deciding so-and-so; this is why we become conscious of (presently) deciding so-and-so.

And the consciousness of a state of affairs *P* being (presently) the case is always somewhat later than the actual fact of *P*'s being the case; it is hardly surprising that the consciousness of making a decision is no exception to this general rule, which is due to the dependence of consciousness on neurophysiology. What is important from the biological point of view is that, in general, the consciousness of something being presently the case does not come too late for the self to react beneficially to the actual fact which is already in the past (which, note, need not preclude its still obtaining at present). For example, it is important from the biological point of view that, usually, the consciousness of several alternative possibilities being open does not come too late for the self to make a decision on which one of them is to be realized, and that, usually, the consciousness of making a decision does not come too late for the self to *revoke* that decision — which decision, indeed, has already been made, but which might still be kept from becoming fully effective.

> *Objector*: I find this interpretation of the result of the Libet-experiment — an interpretation also discussed by other authors (for example, Rosenthal, 2002) — quite problematic for several reasons. First, it is problematic to call something that takes place without the person being aware of it a decision.

Response: But the person *is* aware of its taking place — only a bit later than it is actually taking place, and, as I said, this lag of awareness is quite unavoidable.

Objector: Still, it seems a highly plausible philosophical view about the nature of decisions that it is essential to a decision that the person taking the decision is aware of taking the decision in taking it.

Response: Suppose one insisted on claiming that it is essential to noticing X that the person noticing X is aware of noticing X in noticing X. This view would be quite unhelpful in explaining why we often react to a stimulus X before being able to report that we have noticed X. We should not insist on a philosophical idea if it is unhelpful.

Objector: But you must explain the illusion of the timing of our decisions. When we become aware of our decisions, we certainly do not have the impression of having decided a while earlier but we have the impression of just taking the decision at that very moment. So if it were indeed the case that we make our decisions before being aware of them, we would be quite radically wrong about the moment of our decisions.

Response: According to the Libet-experiment, the brain initiates (the process that leads to) the movement c. 550 msec before the movement and c. 350 msec before the first awareness that one is going to move. Therefore, if the self makes its decision infinitesimally prior to the initiation of movement by the brain — as I have proposed — then the timing of that decision in consciousness is erroneous by c. 350 msec. Do you call *this* 'being quite radically wrong about the moment of our decisions'? I do not think that it is radically wrong from the biological point of view — the point of view that *really* counts. I have already explained above why we become aware of our decisions at all; it remains for me to explain in a different manner than I already did why we become aware of them as being *simultaneous* to our being aware of them, which, indeed, is an illusion (though a biologically benign one, as I have argued). However, let me emphasize first that an explanation of this illusion is required of *any account* of the Libet-experiment, not just of mine. In my view, then, the reason for the said illusion is simply that, due to our neurological organization, we cannot represent *in the conscious present* an unconscious decision of ours that occurred c. 350 msec earlier (than the conscious present) as a — conscious or unconscious? — decision of ours that occurred c. 350 msec *earlier than now.* For representing the decision in question *as being past for such a short time* is biologically unimportant, and therefore evolution made no neurological provisions for having it represented *as being past for such a short time* in consciousness (that is, in *naked consciousness, without* measuring instruments coming to its aid). Since there was no advantage to be gained from the alternative course, evolution took the simplest and most economical course. A most welcome side-effect of the resulting benign illusion is that the unity of the self *as agent* and of the self *as subject of consciousness* is underscored for the self, strengthening the self's

awareness of that very unity (which, indeed, is nothing less than the evolutionary *raison d'être* of the self).

Objector: Be that as it may, but in many cases we do something without any antecedent decision and nonetheless experience our doings as spontaneously brought about by ourselves. You do not seem to have a plausible response to the Libet challenge for these cases.

Response: If we experience our doings as spontaneously brought about *by ourselves*, then we invariably experience them as something we have decided to do. If we experience our doings *not* as something we have decided to do, then we *do not* experience them as spontaneously brought about by ourselves, but as something that is happening to us, spontaneously or not. Note that the experienced decision need not be distinguished in consciousness *as being antecedent* to the also experienced onset of the action; it may also appear — and often does in fact appear — to be simultaneous to the latter. This is the phenomenal situation. Now, what are you trying to tell me? That we sometimes experience certain doings as spontaneously brought about by us without our decision? I have just argued that there are no such experiences. That we sometimes experience certain doings as spontaneously brought about by us, with our decision, but without any *antecedent* decision of ours? As I just said, this often happens, but it is no challenge to my position. That we sometimes experience certain doings as spontaneously brought about by us, with our (experienced) decision, but in fact there is *no* effective decision of ours, antecedent or otherwise? Yes, we can be — and sometimes are — under the *illusion* that we spontaneously bring about certain doings, with our decision — while these doings are, as a matter of fact, entirely due to other factors and while there is no real (effective) decision of ours (see Wegner, 2002). But my point has been that the Libet-experiment does not force us to accept that we are *always* under such an illusion. This is my response to the Libet challenge, covering all cases under that challenge.

5. How the Brain Is a DOMINDAR and Macroscopic Indeterminism

So much for the Libet-experiment. Now, it is a necessary condition of the brain's being a DOMINDAR that there is indetermination in the macroscopic physical world, indetermination which is relevant to the survival of organisms, and enough of it to be detectable. Suppose the brain is in fact a DOMINDAR and there is a lot of biologically relevant indetermination in the macroscopic physical world. How does the brain detect it? And how does the brain transform what it has detected into the consciousness of possibilities of action *now* open to the subject of consciousness and agency?

Nobody, to date, knows the answer to these questions. Nobody, it seems to me, *looks* for an answer to these questions. The reason for this situation is that most researchers regard the macroscopic physical world as evolving deterministically (or *practically* deterministically, if they wish to honour what they believe to be the, as it were, subatomically small probability that quantum indeterminism makes itself felt in the physical macro-world). The fact that much of what happens in the physical macro-world is entirely beyond the pale of predictability does not disturb the usual researchers in their dogmatic slumbers; they have so thoroughly internalized the lesson from chaos theory that unpredictability is no sure sign of indetermination, that they ignore the fact that unpredictability must nevertheless be taken to indicate indetermination with a probability greater than 0.5 — in the absence of contrary evidence (and an *a priori* belief in determinism is *no* such evidence). It also does not disturb them that if determinism is taken to rule in the physical macro-world, then brains must be regarded as incessantly providing their users with ineradicable illusions that have no evolutionary point to them at all.[10]

> *Objector*: If we accept that what we experience is that our decision is undetermined by *all* antecedent factors, then, under determinism, we would indeed labour under an illusion, and an ineradicable one at that. But why would the illusion be pointless? The illusion is necessary to motivate deliberation, and thus the illusion is itself a causal factor in bringing about a rational decision.
>
> *Response*: I have no idea what could be the point of deliberation and rational decision under determinism. Note that determinism implies that what course of action you will now take has forever been fixed (such that the Laplacian Demon could have foretold the course of action you will now take, say, 2000 million years ago). Deliberation and rational decision are, therefore, irrelevant under determinism; if they nevertheless occur, they themselves, far from being 'causal factors' for anything, have forever been determined to occur — even long before the beginning of biological evolution — as inward epiphenomena that absurdly accompany the outward course of events.

[10] In Wegner (2002) a large amount of psychological material is compiled to demonstrate the thorough illusoriness of the experience of free personal authorship of action; but the material is far from inductively establishing the desired conclusion. Moreover, Wegner presupposes a rather primitive Humean model of apparent personal authorship, the model of 'apparent mental causation', which model is far from compelling. Wegner has nothing to say about the evolutionary point of the alleged illusion, which refuses to go away even if it is — allegedly — exposed. It cannot have anything to do with human social life (ibid., p. 342), since the experience of free personal authorship of action would also occur, we may be sure, in an isolated human being.

I have a few speculations to offer as to how the brain is a DOMINDAR. First of all, the brain is a fallible DOMINDAR: not always is there in fact the indetermination in the physical world that the brain tells us there is. Second, the brain does not tell us of *all* the indetermination there is in the environment of our body. We may be sure that some of this indetermination is not noticed by the brain at all, indetermination that is merely at the subatomic level, for instance. But probably there is also *macroscopic* indetermination in the environment of our body (which environment is taken to include, as its limit, the body itself, as I said) that the brain does not notice. From the indetermination the brain notices, it selects the indetermination worth reporting according to *relevancy* (for the survival, or at least the well-being, of the organism) and *restrictability* (since the biological point of detecting and reporting indetermination is to subsequently restrict that indetermination advantageously). I am not saying, however, that all physical indetermination that the brain reports to the self is biologically relevant to the organism and restrictable by the self of the organism. As in other areas of life, we may count on it that there is no perfect fit between a biological faculty and its evolutionary purpose. Sometimes a biological faculty is in error, failing to fulfill its evolutionary purpose, and sometimes it works — 'meaninglessly' — in excess of it.

Finally, the indetermination selected by the brain as worth reporting is classified according to relative importance, so that the self, in consciousness, is ultimately presented with a relatively clear spectrum of weighted alternatives open to it. Then the decision what to do is up to the self.

The crucial question is this: how does the brain manage to notice action-relevant macroscopic physical indetermination? The brain is a macroscopic organ monitoring the rest of the body, the outside of the body, and — least of all — itself. The monitoring is effected via the transmission of physical signals. The brain registers indetermination at a time t in the system that consists of the brain, the rest of the body, and the outside of the body if the totality of the physical signals processed by the brain strikes at t a symmetrical pattern regarding future developments. In such a situation it becomes impossible for the brain to predict how, in certain respects, things will continue to happen. The brain registers this situation as a case of indetermination, and as a case of indetermination in which it is itself involved: in such a manner that the indetermination extends also to a relevant part of its own future activity. The brain may sometimes be wrong about this; for although in some cases the brain cannot predict how things will continue to happen, it is doubtless in some of these cases entirely determined how

things will continue to happen — due to factors that are hidden to the brain. The important thing is that we cannot assume *a priori* that the brain *must always* be wrong when it translates unpredictability as indetermination. More likely than not, the brain is more often than not quite right in making this translation. In support of this position I offer the following evolution-theoretic considerations.

6. An Argument for Macroscopic Indeterminism

Evolution has led to the development of organisms with a monitoring and governing organ: the brain. But if determinism ruled in the physical macro-world, *brains*, we can take it, would never have developed. For what would have been the evolutionary advantage of their developing? If determinism ruled in the physical macro-world, then there would be nothing in that world that needed controlling, and hence nothing would need to be monitored or governed by any organ. For under determinism, everything happens automatically, with absolute precision and with inexorable necessity. Thus, unless there is indetermination of considerable extent in the physical macro-world, the emergence of brains is *absolutely pointless* from the evolutionary point of view. This is true if brains are regarded as entities that, under macroscopic physical indeterminism, would be DOMINDARs. But it is also true if we consider brains merely as highly complex *multi-possibility reactors*, much more complex than other multi-possibility reactors (cars, pianos, computers, etc.) but nevertheless *reactors* of the multi-possibility type, which, if they function well, yield — according to the laws of nature that govern them, without any margin of indetermination — a specific exclusive physical output for each actualized physical input in a set that comprises *several* (in principle) possible physical inputs.[11] Under macroscopic physical determinism, the structural complexity of *every* apparatus, natural or artificial, is *pointless* that makes in advance provision for realizing at a time t one or another of several incompatible alternatives regarding the physical macro-world,[12] where each of these alternatives is possible at time t. Why provide for the realization of one or another among several such alternatives — even if only in such a manner that the realization

[11] Multi-possibility reactors resemble Gary Drescher's *situation-action machines* (in Drescher, 1991). But DOMINDARs are something else than Drescher's *choice machines*. If we compare a multi-possiblity reactor to a piano, then a DOMINDAR might be compared to a piano *plus* a piano-player (where the piano-player may itself be just instrumental for another, higher-order player).

[12] In the car, this provision is manifested by the steering wheel; in the piano, it is manifested by the piano keyboard; in the computer, it is manifested by the computer keyboard.

merely amounts to a law-determined *reaction* to a given physical condition, as in a multi-possibility reactor — *if*, under macroscopic physical determinism, it is true of only *one* thing at any moment in time that it *can* happen in the physical macro-world (namely, the one that does in fact happen)? When evolution ran a course that led, let's suppose, merely by (microscopic) accidental mutation and subsequent natural selection to the development of macroscopic devices that are *geared* for implementing *choices* (made — by the devices themselves or by something else — between at least two incompatible alternatives that are each possible at the time in question), had evolution then forgotten that macroscopic physical determinism is true? Was it ignoring it?

Objector: I have no idea of what the argument is supposed to be here. It seems plain that being governed by a more complex mechanism (being governed by a brain rather than a primitive agglomeration of a few neurons) may be an evolutionary advantage for an organism in providing the capability of more differentiated reactions to information from the environment.

Response: Being governed by a more complex mechanism not only may be an evolutionary advantage for an organism, it is such an advantage in fact. But the question is whether it would be an evolutionary advantage even under determinism. I maintain that, under determinism, being governed by a complex mechanism is no evolutionary advantage for an organism. Hence, under determinism, there is no good reason for such mechanisms to develop. Hence, under determinism, they would not have developed. Now, this argument can be attacked at several points. One objection is rather unlikely to be raised: the objection that even if under determinism there were no evolutionary advantage for an organism in being governed by a complex mechanism, there still would be good reason for such mechanisms to develop — even under determinism. Another objection, which is rather more likely to be raised, is the objection that even if under determinism there were no good reason for complex governing mechanisms to develop, they might nevertheless have developed even under determinism. I address this objection below.

Objector: My objection is neither of the two you just mentioned. I hold that being governed by a complex mechanism is an evolutionary advantage for an organism *even under determinism*.

Response: I have tried to argue the contrary above. Let me try again. Suppose we are playing a rather strange game against each other. The game consists in two players alternately pushing their respective game-buttons: each time a button is pushed a new game-situation appears on the display, replacing the previous one. The game is strange because, unbeknownst to us, each and every stage of it is determined by the rules of the game once its initial situation has been fixed, which is done by throwing dice, say. Unbeknownst to us, the game has, therefore, already

been decided in its very beginning. But, due to lack of knowledge, we are far from just going mechanically through the motions. Moreover, there is a rather striking contrast between us. While I have merely one button for pushing (with the attitude 'well, let's see what comes next and hope that it's not bad'), you have several such buttons and a fancy computer which advises you, in reaction to the game-situation already on display, which button to push in order to make progress towards your goal of winning the game. Thus, you have the capability of more differentiated reactions to information from the environment, so to speak, while I don't have that capability. But are you because of having that capability at an advantage in this game, as compared to me? Evidently, that capability cannot give you an advantage over me, since whatever your computer tells you to do and whatever button you push, the *new* game-situation on the display will be what it has from the beginning been determined to be. It is a predetermined stage on your predetermined way to your predetermined winning — or predetermined losing. Unbeknownst to you, your plurality of buttons and your fancy computer are just so many useless gadgets.

Objector: I am not at all sure whether the story you offer is coherent. How might I fail to notice that my plurality of buttons and my computer are just useless gadgets?

Response: That is a further question, but it can be answered in the following way: With each of your buttons a certain result (a new game-situation) is firmly connected, and you know in each case which one it would be. The computer, in turn, tells you — veridically — at each stage of the game when it's your turn to make a move (i.e., to push a button) that it is rational to push this or that button, with the consequent result. What you don't know, however, is this: that the computer is determined to tell you at each stage of the game, when it's your turn to make a move, to push *that* button which has the result connected to it that is determined to happen anyway at this stage. The — perverse, but possible — situation is simply that the course of the game is both determined from its beginning *and*, with regard to you, as rational as it can be (whether you are predetermined to win or predetermined to lose).

I am of course not saying that the development of the above-mentioned devices for implementing choices is *logically incompatible* with macroscopic physical determinism; for this determinism could, in principle, be of such a kind that the emergence of, say, multi-possibility reactors was itself determined.[13] This would be an *absurd* — that is, an *unnecessarily expensive* — course for nature to take,[14] and therefore a rather unlikely course (even for a complete mechanist regarding nature

[13] In Chapter 2 of Dennett (2003), it is described in detail how this could be.

[14] Against the Dennettian speculations mentioned in the previous note, it should be remembered that nature certainly has no interest in making itself interesting (to *whom?*) by superfluous complexity.

it remains true that nature *normally* follows the course which is the most economical), but it is *not* a logically impossible one.[15]

Therefore, in asserting that if determinism ruled in the physical macro-world brains would never have developed, I am relying on an implicit inference to the best explanation.[16] Made explicit, it is the following inference: Organismic devices geared for implementing choices between several incompatible but possible alternatives with regard to the physical macro-world are widespread throughout natural history, even highly complex devices of this kind, the most prominent examples being brains. The best explanation of this uncontroversial fact is that there are indeed innumerable *choices* that are organismically implemented, that is: that there do indeed exist innumerable organism-dependent realizations of one among several incompatible but at the moment possible alternatives regarding the physical macro-world. And this can only be the case if determinism is, to a considerable extent, *false*, even in the physical macro-world.

Inferences to the best explanation are fallible. But as long as there is no explanation of *the geared-for-implementing-choices fact* which is both better than the explanation that has just been offered *and* preserves macroscopic physical determinism,[17] I prefer to regard the impressive emergence of brains in the course of evolution as an indication of the great extent to which the terrestrial physical macro-world is undetermined (*prior* to additional determination). *Given* this massive macro-indetermination, the unpredictability with which brains are confronted in their monitoring and governing activity must indeed more often than not betoken indetermination.

> *Objector*: But it seems that the set of unpredictable cases for a given brain is a lot larger than the set of cases that might plausibly be assumed to be cases of indetermination. Hence it seems that you are yourself committed to a claim of massive illusion. Moreover, it does not seem to be the case that we interpret every case where we cannot predict what

[15] The basic idea in making the existence of choice-implementing devices compatible with determinism is this: the provisions for implementing choices among several incompatible *synchronic* possibilities might, as a matter of fact, be merely used to *channel* pure event-causal determination under various (synchronically incompatible, but) *diachronically* compatible possible conditions (i.e., conditions whose forever determined realizations may happen to succeed each other in time). It could, in principle, have been a forever-determined fact that devices develop which, while being indeed choice-implementing devices, are never used as such. But what would be the point of their developing? To make what has forever been 'decided' *seem* as if it had not been forever decided? But why?

[16] Compare: an inference to the best explanation is also at the basis of asserting the counterfactual 'If he left the house, the light would not be on.'

[17] It would be quite unwarranted to assume that any explanation that preserves determinism in the physical macro-world *must ipso facto* be better than any explanation that does not.

will happen as a genuine case of indetermination, so you do not seem to do justice to phenomenology.

Response: One must distinguish between the set M1 of all unpredictable cases for a given brain, which is a very large, presumably infinitely large set, and the set M2 of all unpredictable cases for a given brain *that are taken notice of by that brain and are interpreted by it as cases of indetermination*, which is a much smaller and no doubt finite set. Biological evolution has made sure that the further subset M3 of M2, which contains all the *genuine* cases of indetermination in M2, is not too small compared to M2. Massive illusion, therefore, is out of the question, although M3 does certainly not coincide with M2. Regarding your second objection, it suffices to say that my point was *not* that *we* experience cases of unpredictability *for us* which we always interpret as genuine cases of indetermination, but that *our brains* interpret certain cases of unpredictability *for them* as genuine cases of indetermination, which cases *we* therefore experience (veridically or not) as genuine cases of indetermination — and not as mere cases where we cannot predict what will happen (and which we may or may not interpret as genuine cases of indetermination).

7. Two Models of Action-Determination: Chance-Generator and Decision-Maker

Once it is accepted that the brain is often right in translating unpredictability as indetermination, and as indetermination about which something can be done (via the brain), the question arises in what manner it is determined *what will be done*; that is, the question arises in what manner it is determined how the detected indetermination will be restricted. There are two salient models for this. The first model — where the brain is a DOMINDAR *in its own right* — can do without consciousness; it simply consists in this: the brain contains a physical chance generator (that is, a generator of *genuine* physical chance events: physical events without sufficient cause), and determining which alternative to realize from the several realizable alternatives the brain has detected is left to cerebral gambling (and subsequent mechanical cerebral processes), for which procedure consciousness is not essential. The second model — where the brain is a DOMINDAR *instrumentally for something else* — cannot do without consciousness; for, according to it, consciousness is precisely the nonphysical medium in which the several realizable alternatives the brain has detected are presented by the brain to the nonphysical self (under normal conditions, quite faithfully), who then, in the light of consciousness, makes an *at least rudimentarily rational decision* regarding which alternative to realize. This decision may, but need not

necessarily, be preceded by deliberation, and under normal conditions it is quite faithfully put into effect by the brain. It far too often turns out to be *the correct decision* for it to be with any likelihood the result of a mere chance process. The instigation by the self of the brain to go into action in a certain manner is indeed an occurrence of nonphysical causation of the physical without accompanying physical causation.[18] But this occurrence of nonphysical causation of the physical cannot interfere with physical causation and the laws of physics, because it is purely and simply the beginning of the realization of one among several *physical possibilities* — involving brain, rest of the body, and outer environment — that the laws of physics, the entire physical past and therefore the sum total of physical causation could not by themselves exclude from happening.

But what about the, supposedly, big sticks that all physicalists carry even if they speak softly:[19] the principles of causal closure, the allegedly trusty weapons they are quick to wield in order to make the world safe for physicalism?[20] Given that there is macroscopic indetermination in the physical world — indetermination that needs to be restricted somehow, since reality *will* continue in a unique way — it is unwarranted metaphysical dogmatism to believe without reservation in the principles of causal closure of the physical world, be it the strong closure principle, according to which every cause of a physical event must itself be physical, or be it the weak closure principle, according to which every physical event that has a cause at all also has a physical cause.

> *Objector*: I cannot see how indeterminism gives reason to doubt principles of the causal closure of the physical world. Would you care to explain?
>
> *Response*: Suppose we have situations of indetermination in the physical world. Hence at certain points in time — moments of indetermination — the further course of the physical world is not determined by its past. Hence there are physical events — each a part of the (relative) physical future that starts with a moment of indetermination — that have no physical cause. Some of these events may have no cause at all (note,

[18] Remarkably, according to Polanyi (1969, p. 403), 'mechanical effects can be produced without force, merely by selection,' and there is, therefore, 'a possibility for conceiving the action of the mind on the body as exercising no force and transferring no energy of its own. Indeed, since it is the peculiar function of the mind to exercise discrimination, it may not even appear too far-fetched that the mind should exercise power over the body merely by sorting out the random impulses of the ambient thermal agitation. We may bear this possibility in mind whenever referring to autonomous centres of decision.'

[19] Cf. a famous saying by Theodore Roosevelt, referring to diplomacy.

[20] Cf. an equally famous saying by Woodrow Wilson, referring to democracy.

> however, that the old principle of sufficient causation: *every event has a sufficient cause*, once believed to be a quasi-logical truth, does not allow this). But it is quite unwarranted to assume *a priori* that *all of them* have no cause at all. If some of them have a cause, then they that have a cause *have a nonphysical cause* and are thus counter-instances to the principles of the causal closure of the physical world — not only to the strong principle, but also to the weak principle, considering that the events in question are physical events *without physical cause*. For this line of argument, the nature of the nonphysical causes can be left quite unspecified. But if an action — a certain physical event — is prior to its realization not only undetermined by *all antecedent physical* factors, but by *all antecedent* factors (that is, by all antecedent *events*, physical or nonphysical) *and* has a cause, then its nonphysical cause can only be a nonphysical agent.

Now, which of the two afore-mentioned models of action-determination is the correct one? Quite possibly they are realized side by side, each being correct in some cases. In any case, *hard* dualists — i.e., substance dualists who accept agent-causation by the nonphysical self — will insist that the second model is not only feasible, and not only appears to be realized, but is *in fact* realized. The problem for hard dualists is that hardly anybody in the philosophical community nowadays believes this.[21] Another bad reason for this general attitude of disbelief — a reason that I have not yet touched on — is the following:

It is agreed on all sides that a rational decision is not a chance event. But most philosophers these days find it very difficult to distinguish between a rational decision and an event that is causally determined by a complex of desires and beliefs to which the event is, in addition, rationally adapted. In their eyes, what else could a rational decision be but just such an event? But a 'decision' in this widely accepted sense is not a *decision* properly speaking, because it is event-causally determined. One might as well call the turning back of a stone that has been thrown straight up into the air 'a decision', '*its* decision'. Moreover, in view of its event-causal determination, the so-called rationality of a commonly so-called rational decision is merely an irrelevant garnish. A rational decision *properly speaking* is determined only by the *decision-maker*, freely (which implies: in a relevant situation of macroscopic prior indetermination), and *in the light* of his or her desires and beliefs, to which desires and beliefs the decision is rationally adapted by the choice of the rational decision-maker, but which desires and beliefs *do not* cause it.

[21] However, there are new neuroscientific results — specifically, results in cognitive neurobiology regarding tethered *Drosophila* fruit flies — that are decidedly *friendly* (this seems to be the exact term) to the DOMINDAR-hypothesis. See Maye *et al*. (2007).

Objector: Given the immense amount of literature that tries to defend the view that genuine decision is compatible with determinism, I find the tone of this last paragraph a bit irritating.

Response: That tone is itself a sign of irritation on the side of the author.

Objector: Another matter of tone: To my taste, you make it all too obvious — by several polemical passages — that you see yourself as a member of a small minority fighting against the mainstream. I believe that this is rather a disadvantage for the text.

Response: It may well be a disadvantage for the text, given human nature. But take my occasional polemical tone as a — comparatively mild — reaction to the flood of undeserved ridicule and contempt that has been let loose on dualists in recent decades. It helps me to deal with what I consider to be a great philosophical injustice.

References

Chalmers, D.J. (1996), *The Conscious Mind* (New York: Oxford University Press).
Dennett, D.C. (2003), *Freedom Evolves* (London: Allen Lane).
Drescher, G. (1991), *Made-Up Minds: A Constructivist Approach to Artificial Intelligence* (Cambridge, MA: MIT Press).
Dretske, F. (1995), *Naturalizing the Mind* (Cambridge, MA: MIT Press).
Maye, A., Hsieh, C-h., Sugihara, G., Brembs, B. (2007), 'Order in spontaneous behavior', *PLoS ONE*, **2**(5): e443. doi:10.1371/journal.pone.0000443
Meixner, U. (2004), *The Two Sides of Being: A Reassessment of Psycho-Physical Dualism* (Paderborn: Mentis).
Meixner, U. (2006), 'Consciousness and freedom', in *Analytic Philosophy Without Naturalism*, ed. A. Corradini, S. Galvan, & E.J. Lowe (London: Routledge), pp. 183–96.
Newton, I. (1962), *Principia. Volume One. The Motion of Bodies*, translated by A. Motte, revised by F. Cajori (Berkeley and Los Angeles, CA: University of California Press).
Polanyi, M. (1969), *Personal Knowledge: Towards a Post-Critical Philosophy* (London: Routledge).
Rosenthal, D.M. (2002), 'The timing of conscious states', *Consciousness and Cognition*, **11**, pp. 215–20.
Torretti, R. (1999), *The Philosophy of Physics* (Cambridge: Cambridge University Press).
Wegner, D.M. (2002), *The Illusion of Conscious Will* (Cambridge, MA: MIT Press).

Paper received April 2007