

# Mental Causation and the Metaphysics of Causation

Michael Esfeld

Received: 30 June 2006 / Accepted: 12 December 2006 / Published online: 11 August 2007  
© Springer Science+Business Media B.V. 2007

**Abstract** The paper argues for four claims: (1) The problem of mental causation and the argument for its solution in terms of the identity of mental with physical causes are independent of the theory of causation one favours. (2) If one considers our experience of agency as described by folk psychology to be veridical, one is committed to an anti-Humean metaphysics of causation in terms of powers that establish necessary connections. The same goes for functional properties in general. (3) A metaphysics of causation in terms of powers is compatible with physics. (4) If combined with the argument for mental causes being identical with physical causes, that metaphysics leads to a conservative reductionism.

## 1 The Problem of Mental Causation

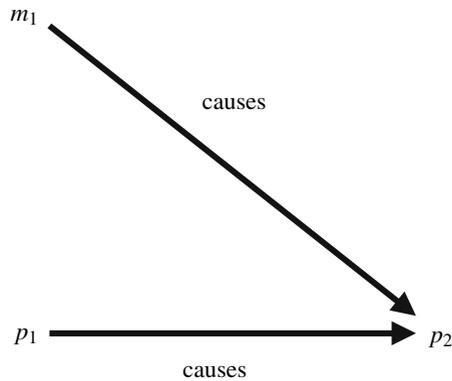
Mental causation is at the centre of the view of ourselves as thinking and acting beings in the world. Mental intentions cause a good deal of our behaviour. Behaviour consists in—or at least includes—physical changes down to changes in the position and the velocity of atoms in the person's body. However, for any physical change there is a physical cause insofar as there is a cause at all. This is what the principle of the completeness of physics tells us, which is firmly grounded on science (for a detailed argument, see Papineau 2002, appendix). Hence, mental causes cannot cause any behaviour that is not at the same time caused by physical causes as well. Suppose that causation is a relation between individual property tokens. Then, for any mental token  $m_1$  that causes a physical token  $p_2$ , there is a physical token  $p_1$  that also causes  $p_2$ . This is the simplest form of the problem of mental causation (Fig. 1).

---

M. Esfeld (✉)

Department of Philosophy, University of Lausanne, Lausanne 1015, Switzerland  
e-mail: Michael-Andreas.Esfeld@unil.ch

**Fig. 1** The simplest form of the problem of mental causation



Taking mental causation and the completeness of physics for granted, there are only two possible ways that one can pursue in order to solve this problem. The one way is to maintain that  $m_1$  is not identical with  $p_1$  so that there is *systematic overdetermination*. Whenever a mental token causes a physical token, the physical token in question also has a physical cause that is not identical with the mental one. This position seems difficult to defend: given that there always are sufficient physical causes anyway (insofar as there are causes at all), why do mental tokens enter as additional causes for some physical effects?

The other way is to maintain that  $m_1$  is identical with  $p_1$  so that there is only one causal relation between  $m_1 = p_1$  and  $p_2$ . If all mental tokens are considered to have a causal impact on behaviour, this position amounts to the view that all mental tokens are identical with physical tokens, more precisely with certain configurations of physical tokens (certain configurations of neurobiological tokens in the brain, for instance). If causes were states or event tokens, they could instantiate different properties. Consequently, one could ask in virtue of which of their properties they cause certain effects. However, if causes are *property* tokens and if there is token identity, it does not make sense to wonder whether the cause brings about the effect qua being physical ( $p_1$ ) or qua being mental ( $m_1$ ). The cause is mental qua being physical, namely qua being a certain highly complex configuration of physical tokens. “Mental” (“ $m_1$ ”) and “physical” (“ $p_1$ ”) are two ways of describing the same property token; their meaning is different, but their extension is the same. I follow John Heil in conceiving property tokens as modes, that is, as particularized ways objects are (see Heil 2003, pp. 137–150). Property tokens in the sense of Heil’s modes are ontologically basic; they cannot be traced back to universals and an instantiation relation between universals and particulars.

Token identity is compatible with the experience of ourselves as thinking and acting beings. That experience is the evidence for mental causation. But that experience cannot settle the issue of whether or not what is experienced, the mental tokens, is identical with physical tokens. This is a theoretical issue to be decided by scientific and philosophical arguments. Given mental causation and the completeness of physics, the problem of mental causation is the main argument for the position that mental tokens are identical with configurations of physical tokens.

The problem of mental causation and the argument for token identity are independent of the stance that one takes in the metaphysics of causation. Consider a Humean regularity view of causation in terms of natural laws. There is an exceptionless, general law couched in physical concepts under which  $p_1$  and  $p_2$  come. More precisely, if  $p_1$  and  $p_2$  are complex configurations of physical tokens, there are exceptionless, general laws couched in physical concepts under which their constituents come, and these laws apply also to  $p_1$  and  $p_2$ . By contrast, if there are laws at all couched in mental concepts, these are *ceteris paribus* laws that admit a lot of exceptions. Thus, the physical laws prevail, and they provide for a satisfactory account of causation that cannot be matched by invoking laws that employ mental concepts. Therefore, the argument against overdetermination applies to the Humean regularity view of causation.

However, the predominance of physical laws is no problem for the identity solution. The causal relation obtains between the property tokens. Since all property tokens are physical and only some configurations of physical tokens are mental tokens, the physical vocabulary, by contrast to the mental one, is the one that we need to express general and exceptionless laws. Nonetheless, if causes are property tokens and if there is token identity, the cause is physical and mental in one. Note that this position is different from the one of Davidson (1970): what is identical are fine-grained entities, namely property tokens, instead of events that may be taken to instantiate mental as well as physical properties. There is no threat of mental *tokens* ending up as epiphenomena, if they are identical with physical *tokens*. Hence, the argument for identity is valid on a Humean regularity view of causation.

The same goes for the view of causation as counterfactual dependence between property tokens (such as the one of Lewis 1973, 2004). The truth value of the propositions expressing the counterfactual dependencies supervenes on the distribution of the property tokens in the actual world. The laws that hold in the actual world are crucial for determining the truth values of the counterfactual propositions in question. For that reason alone, the argument against overdetermination and for identity is valid on a counterfactual account of causation as well (see furthermore Sparber 2005, who argues against Loewer 2001).

If one considers more substantial, non-Humean views of causation, it is clear that they do not admit systematic overdetermination and are therefore committed to the identity solution. If one takes the causal processes in the actual world to be processes of the transmission, the transfer or the exchange of a conserved physical quantity (Salmon 1998; Kistler 1999; Dowe 2000), then it is obvious that (a) effects cannot be overdetermined in a systematic way and that (b) only property tokens that are identical with physical property tokens or their configurations can stand in causal relations. If one traces causal relations back to causal powers, then again it is obvious that there can be no systematic overdetermination of an effect by two powers that are not identical with one another. The theory of causal powers will be discussed in Sect. 3.

Consequently, the argument for the identity solution to the problem of mental causation is independent of the stance that one takes in the metaphysics of causation. It is compatible with all the main positions in the metaphysics of causation. However, this does not imply that all these positions do justice to the

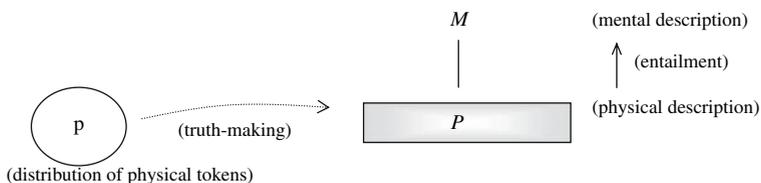
view of ourselves as thinking and acting beings in the world. In the following, I shall claim that this view entails an argument in favour of one particular position in the metaphysics of causation, namely the position of causal powers, and that this position is compatible with physics.

## 2 Agency Against Humean Metaphysics

The identity solution to the problem of mental causation is a sort of ontological reductionism: all there is to the world are physical tokens and their configurations. Only some of these configurations are mental tokens. Consequently, all the entities that there are in the world make true a physical description, and some of them also make true a mental description. Since the entities in question are property tokens, there has to be a systematic relationship between the different descriptions that one and the same configuration of property tokens makes true—in other words, a systematic way from the physical description to the mental one.

The most prominent way of spelling out reductionism is situated within what is today known as Humean metaphysics. Consider David Lewis' thesis of Humean supervenience (Lewis 1986, pp. ix–x). The world is the distribution of physical properties occurring at space-time points over the whole of space-time. These are natural, intrinsic and categorical properties. Their distribution makes true a physical description of the world, and that description entails all the other descriptions: these other descriptions refer to and are made true by certain configurations of physical tokens (Fig. 2).

Combine this view with a Humean stance on causation in terms of certain regularities or certain counterfactuals. Suppose that I have the intention to raise my right arm and that this intention causes my right arm to go up. What does this mean on this view? The description “intention to raise one's right arm” refers to and is made true by a certain configuration of physical tokens, and the description “right arm goes up” refers to and is made true by another configuration of physical tokens. The statement “My intention to raise my right arm causes my right arm to go up” is true on a Humean regularity view of causation if and only if configurations of physical tokens of the first type are spatio-temporally contiguous with configurations of physical tokens of the second type in a regular manner. That is to say, whether or not there is a causal relation between my intention to raise my right arm and my right arm going up depends on what there is elsewhere in space-time. The reason for this dependence is not that the properties of the relata of the causal



**Fig. 2** Lewis' Humean reductionism

relation are extrinsic properties, consisting in certain relations. The reason is that causation depends on laws, and the laws supervene on the distribution of the fundamental intrinsic physical properties over the whole of space-time. If one switches to a counterfactual theory of causation, the account of causation is more sophisticated than a simple regularity view, but the mentioned dependence remains the same, for the truth value of the counterfactual conditionals depends on the laws among other things.

According to Lewis' Humean metaphysics, we have to accept the distribution of the fundamental physical properties over the whole of space-time as primitive (cf. Beebe 2006). Not only that distribution as a whole, but also each property token in that distribution is contingent. That is to say, there are no necessary connections among the property tokens that make up that distribution. In other words, for each single token of a fundamental physical property at a space-time point, it is conceivable and metaphysically possible to hold that token (or its counterpart) fixed and to vary all the other tokens. The physical properties at any given space-time point do not impose any restrictions at all on the physical properties that can occur at other space-time points. Nonetheless, this metaphysics is parsimonious: given the distribution of the fundamental physical properties over the whole of space-time, everything else—including laws and causal relations—supervenes on that distribution. More precisely, if one accepts each fundamental physical property token as primitive, once given the whole distribution of those tokens in space-time, that distribution makes true a complete fundamental physical description of the world. It is possible to derive from that description law statements, causal statements, counterfactual statements and all the other true statements about the world.

However, that metaphysics is at odds with our experience as acting beings (agents) in the world and the conceptualization of this experience in folk psychology. That experience—and its folk psychological description—is veridical if and only if there is a causal relation between mental intentions and behaviour such that the intention brings about the behaviour in the sense that the intention makes it that the behaviour in question exists—given certain favourable background conditions in the body of the person and in the environment. I don't know of any non-circular argument for that claim, that is, an argument which does not presuppose this experience described in these or similar terms. What I shall try to do is to illustrate that claim by contrasting it in two respects with what is the case according to Humean metaphysics.

If the relation between my intention to raise my right arm and my right arm going up is a causal one only in virtue of what is going on elsewhere in space-time, that relation being a causal one does not in any way depend on me. Whether or not configurations of physical tokens making true a statement of the type "intention to raise one's right arm" are regularly followed throughout space-time by configurations of physical tokens making true a statement of the type "right arm going up" is nothing which is under my control. The point is not that being an agent may be an extrinsic property instead of an intrinsic one, depending on, for instance, social relations to other agents. The point is that according to Humean metaphysics, whether there is a causal relation between my intentions and my behaviour depends on there being specific intrinsic properties occurring elsewhere in the world to

which my intentions and my behaviour bear no particular relation apart from similarity. (The argument of Hawthorne (2004) from mental properties against the Humean view of causation amounts to considering these properties as intrinsic; cf. the reply by Weatherson (2004).)

What is more, if the causal relation between mental intentions and behaviour is such that the intention brings about the behaviour, then the relation between the intention—or the configuration of neurobiological tokens with which it is identical—and the behaviour cannot be simply the one of a contiguous spatio-temporal sequence. If that sequence is contingent, the existence of each property token in that sequence is primitive, that is, not grounded on the existence of other property tokens in that sequence. If we acknowledge that the existence of some property tokens in that sequence (those that are identical with the behaviour or with parts of it) is grounded on the existence of certain other property tokens in that sequence (those that are identical with the intention), then we have to endorse necessary connections. In short, if and only if there is a glue that ties the behaviour to the intention such that the intention necessitates the behaviour (given the mentioned background conditions), then we are agents—instead of simply undergoing contingent sequences of mental and behavioural property tokens that happen to satisfy certain regularities in virtue of what is going on elsewhere in space-time.

Consequently, Humean metaphysics cannot admit our experience as acting beings (agents) in the world as being veridical. There is no room for agents in a Hume world. The distribution of the fundamental physical properties over the whole space-time, accepted as primitive, can make true statements of the type “Mental intention *m* causes behaviour *p*”, interpreted along the lines of a Humean theory of causation. But our experience of mental causation in the sense of an experience of agency cannot be accepted as veridical within Humean metaphysics. This objection is important. Mental causation is the argument for mental tokens being identical with physical tokens. However, that argument would be considerably weakened if it turned out that the position which elaborates on that identity is unable to do justice to the way we experience mental causation and conceptualize it in folk psychology. Hence, although the causal argument for mental tokens being identical with configurations of physical tokens is independent of the stance that one takes in the metaphysics of causation, mental causation has a bearing on the metaphysics of causation. We need a more substantial metaphysics of causation than Humean regularities or counterfactual dependence if we take our experience as agents to be veridical.

That conclusion does not have to rely on mental causation. The problem of mental causation is more general than concerning just mental tokens. It applies to all properties in which the special sciences trade. Consider biological properties such as, for instance, the property of being a gene for yellow petals. If the gene causes the plant to have yellow petals in spring, there is physical change down to change in the position and the velocity of some of the atoms in the plant. For that physical change, there also is a physical cause insofar as there is a cause at all. Consequently, there either is overdetermination or identity, and the reasons rehearsed above for taking identity to be the truth of the matter apply.

How can one spell out the claim that, for instance, biological tokens and mental tokens are identical with configurations of physical tokens? According to the most widespread conception, biological and mental properties are functional properties, notably defined by some characteristic effects. One then discovers configurations of physical tokens that are sufficient to cause the effects in question, and the causal argument provides a convincing reason for holding that any biological and any mental token is identical with a configuration of physical tokens.

The trouble Humean metaphysics has with our experience of mental causation extends to functional properties in general. According to Humean metaphysics, the physical properties are categorical only. All there is to the world are the physical properties occurring at space-time points over the whole of space-time. Functional properties, by contrast, cannot be conceived as being categorical only, for what they are consists in the effects that they cause. What these properties are cannot be separated from the causal and the nomological relations in which they stand. Humean metaphysics can acknowledge functional descriptions: the distribution of categorical physical properties at space-time points over the whole of space-time can be such that it makes true certain functional descriptions as derived from the physical descriptions that the physical tokens make true in the first place. But Humean metaphysics cannot admit functional properties: in a Hume world, there cannot be properties whose identity consists in causing certain effects, for all there is in a Hume world is the distribution of categorical properties, and each element in that distribution is contingent. (An analogous point applies, of course, to dispositions: there are no dispositions over and above the categorical properties in the world. But the distribution of the categorical properties over the whole of space-time can be such that it makes certain statements formulated in dispositional vocabulary true, in virtue of making true certain counterfactual conditionals).

To sum up the claim of this section, if our experience as acting beings (agents) is veridical—or, in general, if there are functional properties in the world—then we are committed to a non-Humean metaphysics of causation. If one maintains a Humean metaphysics of causation, then one can provide truth-makers for statements of the type “Mental intention *m* causes behaviour *p*”—and truth-makers for functional descriptions in general—but one cannot accept the way in which we experience mental causation, our experience of agency, as veridical, and one cannot accept the existence of functional properties in general.

### 3 A Metaphysics of Powers

One prominent possibility to move beyond the Humean view of causation is to conceive causation as a physical connection. The most well known theory in that respect is the transference theory of causation. That theory contradicts Hume in maintaining that certain spatio-temporally contiguous sequences of property tokens are causal processes independently of whether or not these sequences occur repeatedly in space-time. Of course, causal processes satisfy law statements, but it is not the law that turns the connection between two or more property tokens into a causal one. Whether or not that connection is causal depends on whether or not the

property tokens in question constitute a process of the transmission, the transfer or the exchange of a conserved physical quantity. These processes are conceived as supervening on the intrinsic physical properties instantiated in the space-time region in which they occur.

Nonetheless, the question is whether the transference theory of causation goes sufficiently beyond Humeanism in order to counter the objection elaborated on in the preceding section. If we take the core idea of what is today known as Humean metaphysics to be the view that all there is to the world is the distribution of categorical physical properties in space-time without there being any necessary connections between these property tokens, then it is possible to receive the transference theory of causation as a version of Humean metaphysics. One can regard the sequences of property tokens that are causal relations as contingent, spatio-temporally contiguous sequences of categorical properties. It is possible to conceive energy and other conserved physical quantities as categorical properties. That is to say, given any one element in any such sequence, it is not necessary that there is any of the other elements. On the contrary, given that as a contingent matter of fact there happen to be sequences of property tokens that make true a description of the type “transfer (or transmission or exchange) of a conserved physical quantity”, there are causal relations in the world. In that perspective, the difference between a transference theory of causation and a theory of causation in terms of Humean regularities or counterfactuals boils down to the following question: can one distinguish certain specific categorical physical properties such that any spatio-temporally contiguous sequence or intersection of sequences of them is by itself causal? Or, the categorical properties be what they may, is it their regular co-instantiation that makes up for the causal relation?

However, the nerve of the objection presented in the preceding section is not the dependence on there being a lot of other sequences of property tokens that are similar to a given sequence for that sequence to be a causal one, but its contingency. The identity condition for functional properties is their causing certain specific effects, the core of our experience of mental causation as conceptualized in folk psychology is its being the ground of the existence of certain bodily movements. If the relation between the causal relata were just a contingent one, these conditions would not be met. For these conditions to be met, there has to be a glue that ties the cause to its effect, other than just the glue of spatio-temporal contiguity. That glue is usually conceived in terms of a power: the cause is a power that brings about or produces its effect in the sense of bringing it into existence, thus necessitating its effect. (Hall (2004) considers two sorts of concepts of causation—counterfactual dependence and production; however, he does not take power theories of causation into account under the latter heading; Kim (forthcoming) also maintains that a Humean theory of causation is not sufficient to take mental causation into account, but limits his consideration of an alternative to the transference theory).

Nevertheless, it is possible to interpret the transference theory as meeting the criterion just mentioned. The physical properties that stand in the causal relations, the conserved quantities such as energy, do not have to be considered as categorical properties only, but they can also be conceived as powers. Accordingly, the transmission, transfer or exchange of a conserved physical quantity does not have to

be regarded simply as a spatio-temporally contiguous sequence of property tokens, but can be seen as the manifestation of a power, namely a power to transmit, transfer or exchange something. Be that as it may, the point at issue is the opposition between a Humean metaphysics of contingency, there being only tokens of categorical properties arranged in a spatio-temporal framework, and a metaphysics of necessary connections, there being powers in the world. The transference theory of causation can be received as fitting into either of these metaphysical positions.

A metaphysics of powers is less parsimonious than Humean metaphysics in that the properties in the world are not regarded as being categorical only, but as being powers. Powers are dispositions that cannot be reduced to a categorical basis. Properties bestow powers on the objects that have these properties or, simply, the properties are powers. I shall use that latter expression for the sake of simplicity (without thereby taking a stance on the issue of the relationship between objects and properties in this paper). There are two versions of an anti-Humean metaphysics of powers discussed in the current literature: the one version considers each property to be categorical and dispositional in one. More precisely, to the extent that there is a distinction between the categorical and the dispositional, it is a distinction among predicates or descriptions instead of properties (see Martin 1997, in particular Sects. 3 and 12); Mumford (1998, chap. 9); Heil (2003, chap. 11). The other version identifies properties with dispositions in the sense of powers (see Shoemaker 1980; Bird 2005, 2007).

The difference between these two versions is not great: the first version does not conceive the distinction between the categorical and the dispositional as an ontological one (so that one cannot even talk in terms of categorical (qualitative) and dispositional aspects of properties). The second version does not conceive powers as pure potentialities, but as real, actual properties—and thus as categorical in a certain sense. Furthermore, each power can be considered as qualitative in a certain sense, since it is the power to produce certain specific effects. In short, properties qua powers are defined by what they do, and what they do is what they are.

The important point is that powers are not additional properties. Qua being a certain way—that is, qua being qualitative—property tokens are the power to produce further property tokens. To put the matter in terms of object talk, qua being a certain way, that is, by having certain qualities, objects have the power to produce other objects, or other properties in objects. The Humean and the anti-Humean agree about the properties that there are in the world. They disagree about the ontology of properties—whether the property tokens are such that they simply succeed one another or whether they are such that each property token brings about other property tokens.

A metaphysics of powers acknowledges as primitive that properties are powers—more precisely, accepts as primitive the fact that qua being a certain way each property token is the power to produce certain other property tokens. On the other hand, such a metaphysics is not committed to recognizing the whole distribution of fundamental physical properties in space-time as primitive. If these properties are powers, the property tokens in a given space-time region may necessitate the property tokens in other space-time regions. Consequently, only the former have to

be accepted as primitive. For instance, if there is such a thing as the state of the world at a given time, the state of the world at a given time may necessitate the state of the world at future times. In the ideal case, one just has to accept an initial state of the world (the “big bang”) as primitive, and the property tokens that make up that state necessitate the subsequent states of the world. In short, in a metaphysics of powers, the elements that have to be accepted as primitive are loaded with a heavier metaphysical baggage than in Humean metaphysics, but one has to accept only a limited amount of these elements as primitive.

#### 4 The Compatibility with Physics

Both the friends of Humean metaphysics and the friends of powers claim support from physics. There is a central objection to Humean metaphysics deriving from physics. The Humean has to be able to conceive the properties in the world independently of the causal relations in which they stand, that is, independently of the effects that they have. According to Humean metaphysics, the causal relations and the laws of nature vary from possible world to possible world, depending on what else there is in a world in which properties of the type *F* occur. Thus, what *F* is neither depends on the effects that the *F*-tokens have nor on the laws in which *F* figures in different possible worlds. The objection now is this one: we have epistemic access to any property *F* only through the effects that its tokens have. In other words, we can conceive *F* only through the effects its tokens have, that is, as the property that has the effects *G*, *H*, *J*, etc. This objection has been put forward by Shoemaker (1980) and motivates his causal theory of properties, which regards properties as powers. This objection has subsequently been further developed notably by reference to fundamental physics. It claims that our physical theories characterize the fundamental physical properties in dispositional rather than categorical intrinsic terms (e.g. Mumford 2006, pp. 475–477).

However, even if one granted that we have epistemic access to the properties in the world only through the effects they have and that we can consequently characterize the properties only in causal terms, this would not imply that the properties are causal powers. They may have an intrinsic essence that we cannot know (Lewis himself endorses this consequence in Lewis 2001). Nonetheless, the mentioned argument is a forceful objection against Humean metaphysics, since Humeanism is supposed to be an empiricist metaphysics, repudiating any entities beyond what can be seized in terms of the empirical sciences.

Consequently, if the fundamental physical properties were categorical and intrinsic, we could not know them insofar as they are categorical and intrinsic. Their categorical and intrinsic character would be hidden from what can be accessed in terms of the empirical sciences. Let us therefore grant that this objection shows that we have no reason to assume that the physical properties are both intrinsic and categorical. Nonetheless, we are not automatically committed to a metaphysics of powers. There remain two philosophical options: the physical properties are intrinsic, but they are dispositional (powers)—this is the option pursued by Shoemaker in his causal theory of properties and the mainstream of the subsequent

metaphysics of powers—or the physical properties are categorical, but they are not intrinsic.

The latter option can be based on our fundamental physical theories. It is by no means mandatory to conceive these theories as characterizing the fundamental physical properties in dispositional rather than categorical terms. There are strong arguments suggesting that instead of intrinsic properties, there are irreducible relations in the domain of fundamental physics. Quantum physics shows that the states of quantum systems are entangled, and entanglement excludes intrinsic properties underlying the relations of entanglement (unless one goes for hidden variables) (see French and Ladyman 2003; Esfeld 2004). As regards space-time as described by general relativity, if one conceives intrinsic essential properties of space-time points, one runs into well known objections, notably the hole argument. There thus are strong arguments for maintaining that the properties of space-time points consist in the metrical relations in which these points stand (see Esfeld and Lam 2007).

Hence, our fundamental physical theories—quantum physics and general relativity—suggest a metaphysics that is known as ontic structural realism. This is the claim that what there is in the world are structures in the sense of irreducible relations rather than intrinsic properties. The relations are concrete, physical relations (such as spatio-temporal relations, or relations of quantum entanglement). Relations do, of course, require relata, that is, objects that stand in the relations. However, according to structural realism, the objects are nothing but that what stands in the relations (e.g. space-time points, quantum particles), without having any intrinsic properties independently of the relations in which they stand.

Nothing hinders to conceive these structures (relations) as being categorical only. The Humean can therefore counter the objection against Humeanism from our epistemic access to the fundamental physical properties by switching from a metaphysics of intrinsic, categorical properties to ontic structural realism, spelling this position out in terms of categorical relations. The switch to a metaphysics of relations instead of intrinsic properties is in any case something that the Humean has to do in order to adapt Humean metaphysics to today's physics. The claim of the Humean then is that in the case of fundamental physics, the causal-dispositional characterization of properties can always be replaced with a structural characterization that conceives the properties as certain relations (structures), but not as causal relations. Spatio-temporal relations as well as the relations of quantum entanglement are not causal relations. Their characterization in structural terms is not inevitably a characterization in terms of powers to produce certain effects.

However, it would be wrong to use this fact in order to claim that a metaphysics of powers is incompatible with current physics (for such a claim cf. Loewer 2001, pp. 322–324; Field 2003, Sect. 1). A metaphysics of powers can be adapted to current physics by conceiving the fundamental physical relations as being powers to produce further relations, that is, in short, as powerful structures (see Esfeld forthcoming). For example, if there are relations of quantum entanglement, one can conceive these relations as including, insofar as they are certain qualitative relations, the power to produce further entanglement as well as the power to bring about processes of the dissolution of entanglement (state reductions). These

processes then lead to classical physical properties having definite numerical values and from there to ever more complex configurations of physical tokens some of which are biological tokens, or mental tokens.

As mentioned in the last section, powers are not additional qualitative physical properties, but the qualitative physical properties or relations that there are conceived as being the power to produce further such qualitative physical properties or relations. Necessary connections do not add anything to the content of physics. If the world is a Humean one or if there are necessary connections in the world, physics is the same in both cases. In other words, physics does not rely upon a characterization of the fundamental physical properties in dispositional terms; such a characterization can be replaced with a structural and purely categorical characterization. But this fact does not mean that physics forces us to dispense with dispositional talk; physics textbooks contain a lot of dispositional descriptions. Nothing hinders to take these descriptions at face value and to regard both the purely categorical and the dispositional descriptions as descriptions of the same entities, namely powerful structures. Thus, the issue between Humean metaphysics and a metaphysics of powers cannot be decided by referring to physics.

The argument for admitting powers stems from our experience of agency and functional properties in general, that is, in short from the special sciences including psychology. If there are powers, there are powers all the way down to the domain of fundamental physics. Properties that are powers, establishing necessary connections, could not supervene on a base of properties that are categorical only. Of course, powers as such are not agents and do not possess intentionality. There is no question here of proposing that the physical is intentional because physical properties are powers. It is only certain highly complex configurations of physical tokens that are mental tokens and possess intentionality. The argument for admitting powers thus is one of coherence: if we look for a coherent view of the world that does not eliminate anything, then we should acknowledge powers in order to do justice to functional properties and the experience of ourselves as acting beings.

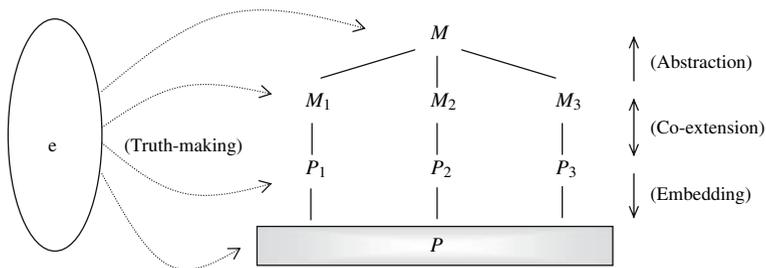
## 5 Conservative Reductionism

Nonetheless, of course, the causal argument for token identity stands—and with it some sort of reductionism. All the powers in the world are identical with physical powers, that is, identical with tokens of fundamental physical powers and their configurations. The power of mental intentions to produce certain behaviour is the power of the configurations of physical tokens with which the intentions are identical (e.g. configurations of neurobiological tokens) to produce other configurations of physical tokens (those with which the behaviour is identical). But if we admit powers on the basis of the argument sketched in this paper, we do not have to spell out reductionism in the austere Humean way explained in Sect. 2.

As already mentioned above, identity is symmetrical: if all biological, mental, etc. tokens are identical with configurations of physical tokens, then some configurations of physical tokens are biological tokens, mental tokens, etc. How can this be so if biological tokens are functional tokens and if some mental tokens

are agency tokens? As argued in this paper, this is possible only if there are powers such that the physical tokens and their configurations are powers. Once we accept that metaphysical view, we are not only in a position to put forward a reductionism (identity theory) that does not provoke the charge of eliminating anything, but we can also maintain that the entities in the world (the property tokens) make true biological descriptions, mental descriptions, etc. in the same way as physical descriptions. There is only one relation of truth-making that applies in the same way to all true descriptions of what there is in the world. We thus get to a conservative reductionism—reductionism, because all the entities (property tokens) in the world make true a physical description, conservative, because some of these entities make true descriptions other than physical ones in the same way as physical ones, and all these descriptions belong to a coherent view of the world including ourselves (Fig. 3).

These descriptions are of course different in meaning. Nonetheless, they are made true all in the same way by the same entities. Either relegating the higher level descriptions to some sort of a secondary status, if not eliminating them at all from a coherent view of the world, or postulating higher level tokens that are distinct from physical tokens because the descriptions differ in meaning is a mistake of what John Heil has branded as the picture theory of language—that is, in short, the view that there is for each meaningful predicate a property out there in the world corresponding to that predicate (Heil 2003, chaps. 2–7; see the papers in Esfeld 2006 for discussion). However, if one leaves behind the picture theory of language, there is no problem in acknowledging that the tokens in the world make true all these descriptions in the same way so that they all belong to a coherent view of the world, provided that one can spell out a systematic relationship between these descriptions. In order to vindicate such a position, one has to acknowledge necessary connections brought about by powers instead of recognizing only categorical properties. That is the manner in which the metaphysics of causation bears on mental causation and the general project of a coherent and comprehensive metaphysics of the world including ourselves.



**Fig. 3** conservative reductionism: “e” stands for the entities (property tokens) in the world, “P” for a fundamental and universal physical theory, “P<sub>1</sub>, P<sub>2</sub>, P<sub>3</sub>” stand for descriptions in the vocabulary of P that are made true by configurations of physical tokens that are identical with mental tokens, “M<sub>1</sub>, M<sub>2</sub>, M<sub>3</sub>” are precise mental descriptions that are co-extensional with P<sub>1</sub>, P<sub>2</sub>, P<sub>3</sub>, “M” is a general, abstract mental description (as regards the possibility to construct such co-extensional descriptions despite multiple realization, see Esfeld and Sachse (2007))

**Acknowledgements** I am grateful to Jens Harbecke, John Heil, Barry Loewer and Georg Sparber for comments on earlier versions of this paper.

## References

- Beebe, H. (2006). Does anything hold the world together? *Synthese*, *149*, 509–533.
- Bird, A. (2005). Laws and essences. *Ratio*, *18*, 437–461.
- Bird, A. (2007). *Nature's metaphysics. Laws and properties*. Oxford: Oxford University Press.
- Davidson, D. (1970). Mental events. In L. Foster & J. W. Swanson (Eds.), *Experience and theory* (pp. 79–101). Amherst: University of Massachusetts Press. (Reprinted in Davidson, D. (1980). *Essays on actions and events* (pp. 207–225). Oxford: Oxford University Press).
- Dowe, P. (2000). *Physical causation*. Cambridge: Cambridge University Press.
- Esfeld, M. (2004). Quantum entanglement and a metaphysics of relations. *Studies in History and Philosophy of Modern Physics*, *35B*, 601–617.
- Esfeld, M. (Ed.). (2006). *John Heil. Symposium on his ontological point of view*. Frankfurt (Main): Ontos-Verlag.
- Esfeld, M. (forthcoming). Structures and powers. In Alisa & P. Bokulich (Eds.), *Scientific structuralism*. Dordrecht: Springer.
- Esfeld, M. & Lam, V. (2007). Moderate structural realism about space-time. Forthcoming in *Synthese*, online prepublication Synthese online <http://www.springerlink.metapress.com/content/1573-0964>
- Esfeld, M., & Sachse, C. (2007). Theory reduction by means of functional sub-types. *International Studies in the Philosophy of Science*, *21*, 1–17.
- Field, H. H. (2003). Causation in a physical world. In M. Loux & D. Zimmerman (Eds.), *The Oxford handbook of metaphysics*. Oxford: Oxford University Press.
- French, S., & Ladyman, J. (2003). Remodelling structural realism: Quantum physics and the metaphysics of structure. *Synthese*, *136*, 31–56.
- Hall, N. (2004). Two concepts of causation. In J. Collins, N. Hall, & L. A. Paul (Eds.), *Causation and counterfactuals*. Cambridge (Massachusetts): MIT Press.
- Hawthorne, J. (2004). Why Humeans are out of their minds. *Noûs*, *38*, 351–358.
- Heil, J. (2003). *From an ontological point of view*. Oxford: Oxford University Press.
- Kim, J. (forthcoming). Causation and mental causation. In *Contemporary debates in philosophy of mind*. Oxford: Blackwell.
- Kistler, M. (1999). *La causalité et les lois de la nature*. Paris: Vrin.
- Lewis, D. (1973). Causation. *Journal of Philosophy*, *70*, 556–567. (Reprinted in Lewis, D. (1986). *Philosophical papers*. (Vol. 2, pp. 159–172). Oxford: Oxford University Press.
- Lewis, D. (1986). *Philosophical papers* (Vol. 2). Oxford: Oxford University Press.
- Lewis, D. (2001). *Ramseyan humility*. Manuscript dated 7 June 2001. (Forthcoming in D. Braddon-Mitchell & R. Nola (Eds.), *The Canberra plan*. Oxford: Oxford University Press).
- Lewis, D. (2004). Causation as influence. In J. Collins, N. Hall, & L. A. Paul (Eds.), *Causation and counterfactuals* (pp. 75–106). Cambridge (Massachusetts): MIT Press.
- Loewer, B. (2001). Review of Jaegwon Kim, mind in a physical world. An essay on the mind-body problem and mental causation. *Journal of Philosophy*, *98*, 315–324.
- Martin, C. B. (1997). On the need for properties: The road to Pythagoreanism and back. *Synthese*, *112*, 193–231.
- Mumford, S. (1998). *Dispositions*. Oxford: Oxford University Press.
- Mumford, S. (2006). The ungrounded argument. *Synthese*, *149*, 471–489.
- Papineau, D. (2002). *Thinking about consciousness*. Oxford: Oxford University Press.
- Salmon, W. C. (1998). *Causality and explanation*. Oxford: Oxford University Press.
- Shoemaker, S. (1980). Causality and properties. In P. van Inwagen (Ed.), *Time and cause* (pp. 109–135). Dordrecht: Reidel. (Reprinted in Shoemaker, S. (1984). *Identity, cause, and mind. Philosophical essays* (pp. 206–233). Cambridge: Cambridge University Press.
- Sparber, G. (2005). Counterfactual overdetermination vs. the causal exclusion problem. *History and Philosophy of the Life Sciences*, *27*, 479–490.
- Weatherson, B. (2004). Humeans aren't out of their minds, unless everyone is. Retrieved <http://www.brian.weatherson.net/papers.html>