

Max Kistler

The Mental, the Macroscopic, and their Effects

Epistemologia (Genova, Italie), 29 (2006), p. 79-102.

Abstract

In questo articolo viene proposta una concezione del ruolo causale delle proprietà mentali diversa tanto dal dualismo quanto dalle due varianti del materialismo, quella eliminativa e quella riduttiva. Gli eventi che coinvolgono proprietà cognitive sono concepiti in modo da avere anche proprietà fisiche e neurofisiologiche. L'idea è che le proprietà mentali di questi eventi possono avere, nelle interazioni causali, un impatto reale anche sugli eventi fisici: il fatto che l'evento *c* ha la proprietà mentale *F* può essere causalmente responsabile del fatto, mentale o fisico, che l'evento *e* ha la proprietà *G*. L'evento *c* ha responsabilità causale in virtù: 1) di una relazione causale tra gli eventi *c* ed *e* e 2) di una legge di natura che si applica a questi eventi e rende *G* dipendente da *F*. In questo quadro si può dimostrare che l'efficacia causale della proprietà mentale *F* non è vanificata (*pre-empted*) dall'efficacia causale delle proprietà fisiche dell'evento *c* (che è un *F*), contro l'argomento dell'esclusione causale di Kim secondo cui l'efficacia causale verso il basso (*downward causation*) delle proprietà mentali non può coesistere con le relazioni causali fisiche.

La conclusione di Kim può essere evitata distinguendo tra determinazione causale e determinazione *non causale* delle proprietà mentali da parte delle proprietà fisiche sottostanti. Affinché una proprietà mentale di un evento possa essere efficace nel senso di contribuire a determinare le proprietà degli effetti di tale evento, deve esistere una legge psicologica che la coinvolge.

L'efficacia della proprietà cognitiva non dipende dalla possibilità di ridurla alla neurofisiologia: la riduzione di una proprietà mentale equivale alla spiegazione nomologica della sua determinazione non causale da parte di proprietà neurofisiologiche. Si considerino, per esempio, due casi di leggi psicologiche, l'una che coinvolge rappresentazioni della posizione spaziale di un agente, l'altra le associazioni tra stimoli e risposte apprese durante il condizionamento. Queste leggi sostengono l'efficacia causale delle proprietà cognitive almeno nei casi in cui non c'è nessuna legge corrispondente ai livelli inferiori, e in particolare al livello neurofisiologico (come probabilmente avviene nel caso del cervello). Si può spiegare l'assenza di leggi siffatte col fatto che, al livello di descrizione neurofisiologico, l'evoluzione degli eventi cerebrali è caotica.

Mental properties are macroscopic properties of cognitive systems. Such properties are naturally determined by the physical, in particular neurophysiological, properties of the parts of these systems and their nomological interactions. This is compatible with psycho-physical reduction by the discovery of the nature of that determination relation; yet reduction does not lead to the identification of the mental property with any underlying physical property. Their implication in psychological laws guarantees the causal efficacy of mental properties. This efficacy may not be paralleled by the efficacy of physical micro-based macroproperties because the evolution of the brain is chaotic on a microscopic scale.

It has been argued¹ that there are only two consistent ways to make sense of the intuition that the mind interacts causally with the physical world. Either one embraces dualism, holding that some causes are non-physical, in particular, persons, or one accepts materialism, holding that only physical properties possess real and non-derivative causal powers, whereas mental properties can

¹ See Kim (1998).

have at most derivative causal powers. However, the argument continues, anti-reductive materialism, holding that mental properties are irreducible to physical properties but nevertheless causally efficacious, is an unstable position that can be shown to collapse either into reductive materialism or into dualism. According to reductive materialism, only those properties can be acknowledged as real and causally efficacious that are reducible to physical properties. With respect to mental properties, future cognitive science will show whether they can be reduced to neurophysiological properties or not. This makes two materialist positions conceivable: reductive materialism (or type physicalism) if mental properties turn out to be reducible, and eliminativism if they do not. However, a physicalist who wants to avoid dualism, cannot hold that mental properties are irreducible and nevertheless causally efficacious. As Kim says, “a middle-of-the-road position [...] is not available. More specifically, [...] a physicalist has only two genuine options, eliminativism and reductionism”. (Kim 1989, p. 267)²

In this paper, I would like to suggest that anti-reductive materialism is not the only remaining alternative steering a middle course between the scylla of dualism and the charybdis of reductive materialism (or eliminativism). It is not necessary, in order to avoid both type reductionism and eliminativism, to find an argument for the impossibility in principle of reducing mental properties to physical properties. It is an open question, and up to science not philosophy, to find out whether reductions, local or global, of mental properties can be achieved, maybe by constructing new mental or neurophysiological concepts. The issues of reducibility and causal efficacy are independent. The causal efficacy of a mental property is neither threatened by a reduction (as both dualists and some reductionists claim) nor is a reduction required to secure it (as some physicalists think). Of course, the reducibility of a property plays an important role for our *understanding* its causal efficacy. The reduction of a cognitive capacity, for example by discovering a mechanism underlying its exercise³, is the best way to explain why and how that capacity has a causal influence on the world. But its efficacy itself does not depend on its reducibility but rather on the laws of nature it is embedded in, which link it to other properties at its own level.

My plan is the following. First, I sketch a framework in which the question of how mental properties are causally efficacious can be addressed without prejudging it either in the direction of eliminativism or type physicalism or in the direction of dualism. Then I try to show that this framework allows us to defend the possibility of genuine mental causation against Kim’s argument that mental properties can have only derivative causal efficacy.

One preliminary remark: I shall here ignore the problem of the intentional character of mental states and properties. At least some mental states have content, and on an externalist construal of content, this makes their identity determined by things which are too far away from the time and place where the mental state is localised to be able to contribute to its causal efficacy. Thus we face

² Kim speaks of physicalism instead of materialism. Physicalism is the doctrine that 1) every *object* is composed exclusively of parts which have only physical properties, and that 2) every intrinsic *property* of a complex object can in principle be reduced to the physical properties of its parts. The position I shall defend and which I call “materialism”, accepts the first thesis of physicalism but considers the truth of the second part as an open, and empirical, question.

³ On the notion of a mechanism, cf. Glennan (1996) and Machamer et al. (2000).

the puzzle of how mental states can be efficacious in virtue of their content. However, this problem can be treated independently⁴ from the problem addressed here: How can we understand the causal efficacy of global properties of complex objects, without being forced to accept either dualism or eliminativism or type reductionism? My strategy is to put this question within a framework which is more general than the issue of mental properties. Part of the problem of mental causation is just as much a problem for the theory of macroscopic causation in general⁵; clarification of the nature of the latter may therefore help making progress with the specific issue of mental causation.

Causality, causal responsibility, and causal explanation

Take an everyday statement of mental causation. The thought that the noise from the street disturbs my concentration causes my decision to close the window. Each of these mental events, of which the statement says that they are related as cause and effect, also has physical properties⁶. Or take a case of psychophysical causation where a mental event causes a physical event. My decision (a mental event) causes my closing the window (understood as a bodily movement: a physical event). The question quite naturally arises, whether the mental property of being a thought that the noise from the street disturbs my concentration makes a causal contribution to my decision to close the window, and whether the mental property of being a decision to close the window makes a causal contribution to my actually closing it. More precisely, the controversial issue is whether the efficacy of these mental properties is or is not precluded (or “preempted”) by the efficacy of the underlying physical or neurophysiological properties of those events.

A satisfactory reply to this question must build on a more general analysis of the contribution different properties of an event make to what it causes. A theory of causation must provide a conceptual framework making it possible to say what it is about a cause that is responsible for its effect having specific properties. If a red billiard ball causes a white billiard ball to move off in a given direction with a given speed, it causes this movement in virtue of its momentum (among other factors), but not in virtue of its colour.

⁴ See Dretske (1988), Lewis (1994), Braddon-Mitchell and Jackson (1996), Chalmers (1996). My justification for sidestepping the issue of the causal role of content is that it is an instance of the general problem of finding what Crane and Mellor (1990, p. 194) call a “local causal surrogate” *S* for a property *P*, where *P* is a relational or otherwise extrinsic property, and *S* is an intrinsic property which is *directly* (and locally) causally responsible for *P*’s effects. Postulating an intrinsic mental property directly responsible for the effects of a subject’s thoughts (which are themselves extrinsic, because their nature is determined by their relations to states of affairs localised outside their thinker), can be seen as analogous to postulating a local electrical field which is directly causally responsible for the acceleration of an electrical charge *C*₁. The local field strength is the local causal surrogate for the relational (extrinsic) property of *C*₁ to be at a certain distance from another charge *C*₂.

⁵ This is often observed; see, e.g., Baker (1993, p. 79; 1998, p. 261).

⁶ My adoption of this Davidsonian way of speaking about events is not meant to prejudge the question of the relation between the physical and mental properties of these events, which is the topic of the paper. At this point, the question could just as well be formulated in the alternative terminology proposed by Kim (1973), according to which an event is an object’s having a property at a time. In this terminology, the event of my thinking that the noise from the street disturbs my concentration is accompanied by (or supervenes on) an event of my (or my brain’s) having some neurophysiological property.

Referring to fact F_1 (the fact that the red ball hit the white ball with momentum M) causally *explains* (together with the fact that the white ball is at rest and no other factors interfere) fact F_2 (the fact that the white ball moves with M after the shock). This explanatory relation has an objective truthmaker. We can express the truth-maker by saying that F_1 is *causally responsible* (CR) for F_2 ⁷. This accounts for the fact that the distinction between F_1 (the fact that the red ball hit the white ball with momentum M) and F_3 (the fact that the red ball was red when it hit the white ball) reflects an objective difference in causal influence with respect to the movement of the white ball. This difference does not depend on an explanatory difference; F_1 but not F_3 would still be causally responsible for the effect even if there were no explanation-seekers or explainers at all.⁸ The concept of causal responsibility allows us to express the fact that some properties (in our case, momentum) of the cause event make an objective difference to what the effect is like, and others (in our case, colour) do not.

(CR) The fact that c is F is *causally responsible* for the fact that e is G if and only if c is a cause of e (at the level of particular events), c exemplifies F , e exemplifies G , and there is a law (which is in general non-strict) according to which instantiations of F tend to produce instantiations of G .

In (CR)⁹, “ c is a cause of e ” designates a causal relation between particular events, which may be grounded in the transference of an amount of some conserved quantity, such as energy¹⁰. An event is here construed as a particular whose identity conditions are given by the limits of the space-time-zone it fills. The statement that the fact¹¹ that c is F is responsible for the fact that e is G , is meant to imply both that c is a cause of e and that c 's property F is causally efficacious in producing or bringing about an event with property G .

In order to guarantee that causal responsibility (CR), which depends on properties, is as local as the causal relation between events, we need to conceive of the constitutive properties F and G of the facts that c is F and that e is G , as property-instances or tropes, not universals¹². Only a spatio-temporal entity such as a trope can be locally efficacious, whereas universals are not located in

⁷ For the notion of causal responsibility, see Kistler (1999a); (1999b).

⁸ Putnam (1992, p. 47ff.) and Hardcastle (1998) suggest to reduce the *ontological* problem of which properties are causally efficacious to the *epistemological* problem about which explanations are pragmatically preferable. This turns the direction of philosophical explanation on its head: some explanations are objectively more *correct* than others, given the explanandum but independently of our interests. The momentum, not the colour is efficacious with respect to the momentum of the recoil ball.

⁹ Causal responsibility is similar to what others call “qua-causation” or “quausion”. Cf. Horgan (1989), McLaughlin (1993), Kim (1993a), Marras (1998).

¹⁰ This is defended in Kistler (1998; 1999a).

¹¹ Given that many events are time-slices of objects, what I call “facts” may be considered as a category containing what Kim (1973) calls “events” as a special case.

¹² This has been proposed specifically for mental causation by Macdonald and Macdonald (1986, p. 37-40), Heil (1992, p. 136-139), Ehring (1997) and Robb (1997), and for causation in general by Keith Campbell. In his words, “the terms of every real causal sequence are one and all of them particulars. When you drop it, it is the weight of this particular brick, not bricks or weights in general, which breaks the bone in your particular left big toe.” (Campbell 1990, p. 113).

space and time at all¹³. However, this does not by itself solve the problem, for the crucial question is how to identify the causally efficacious tropes. The following nomological criterion ties the identity of a property (trope) to its lawful relations to other properties (tropes of other types).

(Nomological criterion of property identity) Property P is identical to property Q if and only if, for all properties R , P stands in a lawful relation N to R if and only if Q stands in the same lawful relation N to R .¹⁴

$$\forall P \forall Q \{P=Q \leftrightarrow \forall R \forall N [N(P,R) \leftrightarrow N(Q,R)]\}$$

P and Q are identical properties if and only if they share all their nomological dependency relations to other properties and, as a consequence (because causal responsibility is determined by lawful dependency), all their causal responsibilities.

Consider once more the example of the red billiard ball causing a white billiard ball to move in a given direction with a given speed. There is a law linking the momenta and thus the speeds of these balls, but no law linking the colours to the speeds. Therefore, the colour trope of the red billiard ball is different from its speed trope, and the causal responsibility for the speed of the white ball after the shock lies with the speed not the colour of the red ball.

On this criterion, mental and physical properties come out as different: They are embedded in different laws. In this framework, the crucial question becomes whether there exist psychological laws between different mental properties. This is an empirical question¹⁵. If such laws exist then mental properties are causally efficacious.

Preemption, causal-explanatory exclusion, and downward causation

The thesis of the causal efficacy of global properties of complex macroscopic entities such as mental properties of human beings must face the following challenge. Let us accept the idea that mental properties (or their local surrogates) are intrinsic global properties of people and other cognitive systems, and that global properties of macroscopic objects O are exclusively determined by the physical properties of the parts of O , and their interactions. One way to spell this determination relation out is to say that global properties *strongly supervene* on the properties of the parts¹⁶. The relevant concept to be used here is strong mereological supervenience of the properties of wholes on the properties of their parts:

Necessarily, if any macroscopic object with parts P_1, \dots, P_n , has some property A , there exist properties $B_1(P_1), \dots, B_n(P_n)$ of its parts and relations between the parts $R_1(P_1, \dots, P_n), \dots$ such that necessarily every macroscopic object that has parts P_1, \dots, P_n , with properties $B_1(P_1), \dots, B_n(P_n)$ and relations $R_1(P_1, \dots, P_n), \dots$ has property A .

¹³ Of course, universals are usually said to be “wholly present” at each of their instances, and the instances *are* located in space and time. At any rate it is the instance that acts at a particular place and time.

¹⁴ Achinstein (1974) has proposed a similar criterion of property identity in which causal equivalence plays the role here assigned to nomological equivalence.

¹⁵ I shall mention such laws below.

¹⁶ Cf. Kim (1978), Kim (1988b), Kim (1989), (Kim 1993b, p. 168).

In these terms, the problem of mental causation appears as a special case of the more general problem of understanding whether and how different properties of the same event cooperate or compete in causing other events. More specifically, the conviction that the mental is causally efficacious faces the challenge of causal *preemption*: the causal powers of macroscopic properties of an object *O* might turn out to be at best only derivative and at worst inexistent, with respect of the causal powers of the underlying physical properties of *O*'s parts¹⁷.

Kim (1993a) has clearly stated the challenge of preemption against the claim of the causal efficacy of mental properties. Suppose an instance of a mental property *M* causally brings about an instance of a different mental property *M**¹⁸. According to strong supervenience, whenever an organism has *M*, there is a property *P* such that everything that has *P* necessarily has *M*. Then (the instance of) *M* supervenes on (an instance of) *P*, and *M** supervenes on *P**, as illustrated in fig. 1.

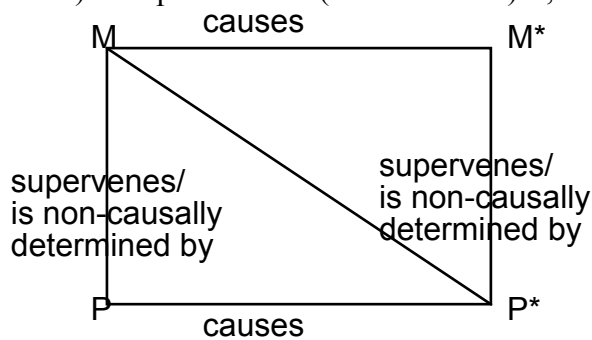


Fig 1. Distinguishing causal and non-causal determination

The crucial question is, what is causally responsible for this instance of *M**, say my decision to close the window? This decision supervenes on a certain pattern of neural activity that will in general be distributed over large areas of the brain. By hypothesis, *M** is not brought about directly in a physical way, e.g. by stimulating a particular neuron, but by a mental cause *M*: by the thought that the noise from the street disturbs my concentration.

Kim (1993a) argues that mental-to-mental causation implies downward causation¹⁹, and therefore violates the principle of causal-explanatory exclusion and the principle of the causal closure of the physical domain. The former says that "there can be no more than a single complete and independent explanation of any one event" (Kim 1988a, p. 233), the latter says that "if you pick any physical event and trace its causal ancestry or posterity, that will never take you outside the

¹⁷ Technically, Kim considers preemption of the mental properties of an individual by physical properties of the same individual, not by properties of its parts. However, it is equivalent to compare the mental properties of *O* with the physical properties of its parts, and with a "micro-based" property of *O*. See below. Cf. Kim (1988b, p. 124). Such a micro-based property (or, in Armstrong's terms, "structural property") belongs technically to the whole, but only thanks to a logical construction. Its only ontological basis lies in the properties of the parts. Therefore, the only sense in which one can say that a person has physical properties is the sense in which it is technically possible to attribute the physical properties of its parts to the person itself.

¹⁸ For simplicity of exposition, I shall occasionally adopt Kim's way of speaking of property instances as causes and effects.

¹⁹ In his defence of downward causation, Sperry (1976) does not resolve the tension arising from his combination of the thesis that microproperties determine macroproperties with the thesis that these macroproperties make an independent causal difference to the very microproperties they are determined by. For doubts about the coherence of such circular determination, see Klee (1984), Stephan (1992, p. 42ff.), Kim (1992).

physical domain. That is, no causal chain will cross the boundary between the physical and the non-physical” (Kim 1997, p. 282)²⁰. First, Kim argues we can make sense of M 's causing M^* only by supposing that M causes M^* by causing P^* , which is a case of downward causation. Second, he argues that we cannot make sense of M 's causing P^* because M 's causing P^* is pre-empted by its supervenience basis P 's causing P^* . He concludes that mental properties are epiphenomenal: they are not themselves causally responsible for anything.

Let us look at the first part of the argument. In the situation sketched in fig. 1, there are two correct answers to the question: What is responsible for M^* ? According to one, it is M ; according to the other, it is P^* . Now Kim argues that there are only three ways in which these answers about why M^* is there can both be true. The first two being unacceptable he concludes that the third must be correct.

1. M and P^* are jointly responsible for M^* .
2. M and P^* overdetermine M^* .
3. M causes M^* by causing P^* .

Hypothesis (1) says that M and P^* are jointly responsible for M^* , although neither would be a complete cause of M^* in itself. Kim argues that (1) is wrong for supervenience tells us that P^* alone is sufficient for M^* . Against hypothesis (2) that M and P^* overdetermine M^* , in the sense that each is a complete cause of M^* , he argues that "overdetermination" implies that there are two "distinct and independent origins" (Kim 1993a, p. 205) of M^* . As the case is described, M and P^* are indeed not independent in the way in which two same-level causes overdetermine an effect, as when two independently shot bullets cause the same death. Therefore, in virtue of the principle of causal-explanatory exclusion, M cannot be a further complete cause of M^* , over and above P^* .

The weak point of the argument is that hypotheses (1) to (3) do not exhaust possibilities. Kim conflates "determining", "being a sufficient condition for", and "causally bringing about"²¹. These are very different relations. The argument against (2) makes it clear that he intends

²⁰ Cf. also Kim (1993a, p. 209). To put it more abstractly, "at every time at which a physical state has a cause, it has a fully sufficient physical cause" (Lowe 2000, p. 27). The principles of causal-explanatory exclusion and of causal closure are not equivalent. By treating all explanations of all properties of things which are exclusively constituted by physical parts, as physical explanations, Baker (1998, p. 262/3) gets the result that mental explanations are always physical explanations, and therefore respect the causal closure of the physical domain. However, this move provides no guarantee against the preemption of mental causal explanations by neurophysiological or still more micro-physical explanations, in virtue of the principle of causal-explanatory exclusion.

²¹ Similar points are made by Thomasson (1998) and Jacob (2002). Slors (1998) proposes a more charitable interpretation of Kim's argument. He asks how it is possible to construe the relations of realization (between P^* and M^*) and of causation (between M and M^*) in such a way that they become sufficiently "similar" to make sense of Kim's claim that M and P^* compete for being a "sufficient condition" for M^* , and of Kim's proposed solution in which these relations are concatenated in a chain of sufficiency $M - P^* - M^*$. His result is that this makes sense only if both realization and causation are understood as nomological relations. Given standard interpretations of what realization and causation are, this seems indeed the "largest common denominator", their largest common conceptual element. However, Slors accepts the idea that even under this interpretation of the two relations, M and P^* still compete for nomological sufficiency, and concludes that mental causation can only be saved from being epiphenomenal by either denying that causation or that realization is nomological. However, as I argue in the text, these relations of nomological sufficiency, being of very different types, do not compete; therefore to distinguish them suffices to dissolve the problem posed by Kim. Hence Slors' solution to deny the nomological character of realization is unnecessarily strong.

"overdetermination" to mean "causal overdetermination", as in the case of two bullets causing one death. It is indeed implausible to hold that a mental cause M and its neural subvenient property P overdetermine their effects in this way²². But this is not the issue we are addressing. M and P^* cannot be two parallel overdetermining causes of M^* because the relation of P^* to M^* is not causal: first, supervenient properties and their bases are instantiated by the same object at the same moment, but causal relations require spatio-temporally separate relata²³. Second, strong supervenience is necessary whereas causation is contingent.

In fact, M and P^* both determine M^* but not in the same way: M causes it, in virtue of a lawful link between them, but P^* determines it in a constitutional, non causal way. This shows a plausible fourth way, not considered by Kim, to reconcile the two explanations of the presence of M^* : one is a causal, the other a constitutional explanation²⁴. M and P^* do not causally overdetermine M^* although each is on its own "sufficient for" M^* . "Sufficient for" can have two meanings: M is causally sufficient for M^* and P^* is constitutionally sufficient for M^* .

Here are two simple examples where explanations of these kinds do not compete. Why is there an equilateral triangle on the piece of paper on my desk? First, causal answer: because I drew it. Second, non-causal answer: because there is an equiangular triangle on the paper, and necessarily, all equiangular triangles are equilateral. Why is this gas at temperature $T=50^\circ\text{C}$? First, causal answer: because I just raised its temperature by heating its container. Second, non-causal, answer: because its pressure and volume are P and V , and because it is approximately an ideal gas, and T is proportional to P and V according to the ideal gas law.

Whether this is compatible with Kim's "principle of explanatory exclusion", which says that "there can be no more than a single complete and independent explanation of any one event" (Kim 1988a, p. 233), depends, of course, on how "complete" and "independent" are understood. On one plausible construal, "complete" means "sufficient on its own", and "independent, "not linked by conceptual, logical or metaphysical necessity". Then the account is incompatible with Kim's principle because causal and non-causal explanations of the same fact can coexist although they are independent and each is complete.²⁵

Downward causation

²² Crane (1995, p. 232) and Loewer (2002) observe that overdetermination is implausible only insofar as it implies massive coincidence. However, the implausibility of the "coincidence" kind of overdetermination does not establish the implausibility of systematic overdetermination by mental and physical properties that are nomically related by psychophysical laws.

²³ Kim recognizes the Humean requirement on causation, that cause and effect must be temporally distinct. Cf. Kim (1993a, p. 207) and (1998, p. 44).

²⁴ Cummins (1983) draws a similar distinction between causal and non-causal explanations in psychology. Charles (1992) has advanced the idea that mental properties exist by virtue of composition laws linking them to neurological properties. On Charles' account, these laws are irreducible, similar to the "transordinal laws" of British emergentism (Broad 1925). However, there are no good conceptual reasons against the reducibility of such psychophysical laws of composition. At any rate, the discovery of laws of composition would seem to provide the basis for a reduction of the mental properties involved, though not for their identification with their compositional basis.

²⁵ It would not refute a principle of *causal*-explanatory exclusion because one of these explanations is not causal.

We have undermined Kim's argument that mental-to-mental causation requires downward causation. It is not true that one can make sense of both M and P^* 's determining M^* *only* by supposing that M causes M^* through a downward causal process running from M through P^* to M^* . Nevertheless, we must address Kim's argument that mental properties cannot have any *physical* effects.

What determines or brings about P^* ? There are two candidates: M (through the diagonal, downward path in fig. 1) and P . According to Kim, the only alternatives are

1. M and P are jointly responsible for P^*
2. M and P overdetermine P^* .
3. P causes P^* by causing M .
4. P causes P^* directly, without any intervention by M .

Then he argues that the first three must be rejected. This means that there is no downward causation after all, for only the first three scenarios make reference to downward causation.

According to Kim,

(1) is wrong because P is alone sufficient for P^* . This follows from the principle of the causal closure of the physical. It says that, P^* having a cause at the time of M and P , it must have a purely physical sufficient cause at that time. Therefore, M cannot be a necessary part of the cause of P^* , and P alone causes P^* .

(2) is wrong because it would have the absurd consequence that there is systematic causal overdetermination of brain events by mental and brain events. Moreover, this possibility would violate the principle of causal-explanatory exclusion, according to which there is only one complete independent causal explanation of P^* .

(3) must be rejected first because it is unnecessarily complicated, compared to the hypothesis (4) that P alone causes P^* , and second, because there cannot be a causal chain $P - M - P^*$ due to the simultaneity of P and M .

I shall challenge Kim's argument against (1). P , the physical basis of the mental property M possessed by an individual A , can be construed as a "conjunctive property", which can be expressed by a long conjunction of properties of A 's parts, say descriptions of the states of a large number of nerve cells and synapses. In Kim's terminology, P is a "micro-based" macroproperty : a property that "belongs to a whole in virtue of facts about its parts" (Kim 1988b, p. 142/1993c, p. 124). In this characterization, the "in virtue of" relation is purely logical: " P is a *micro-based property* just in case P is the property of having proper parts, a_1, a_2, \dots, a_n , such that $P_1(a_1), P_2(a_2), \dots, P_n(a_n)$, and $R(a_1, \dots, a_n)$ " (Kim 1997, p. 292). The subject's possessing P is *logically* determined, in virtue of the conjunction operation, by the properties of the subject's parts, and the spatial relations between these parts. By contrast, the instance of mental property M is a global property of A that is *naturally* determined, in virtue of a law of composition, by interactions among neural networks and other

parts of the body²⁶. Both P and M are determined by the properties of A 's parts, and therefore micro-based properties in Kim's sense; however, there is a crucial difference in the way they are so determined: The physical property P is *logically* determined by the properties of the parts and their spatial arrangement, whereas the mental property M is *naturally* determined by laws of composition.

At the level of events, there is just one causal relation between the P - M event and the P^* - M^* event. The issue of downward causation arises only with respect to causal responsibility. My point is that it is empirically possible that M has at least partial *downward causal responsibility* for the fact that the effect is P^* . Let me compare this case with a case in which an incident billiard ball carrying momentum m and fresh red paint on it sets a recoil ball in motion, thereby transmitting a spot of red paint to it. There is only one causal relation between the events of the first ball hitting the recoil ball and the event of the recoil ball moving off with a red spot on it. However, if we distinguish between causation and causal responsibility, we can say that different facts about the cause are causally responsible for different facts about the effect. The momentum of the incident ball is causally responsible for the momentum of the recoil ball after collision, but its being freshly painted red is causally responsible for the recoil ball's having a red spot. Causal responsibility is the ontological ground of causal explanation. Such explanations do not exclude one another, because several such relations can run parallel between facts involving different properties of one given cause-effect pair. Once we have established that the P - M event has caused the P^* - M^* event, we may ask, for each of the properties of the effect event, which of the properties of the cause event is causally responsible for it. The existence of a psychological law linking M to M^* justifies attributing to M the causal responsibility for the fact that the effect has property M^* .

Causation between mind-brain-events differs from the billiard balls case²⁷, because a mental property M^* such as a decision to close the window is not independent from the underlying conjunctive neural property P^* , in the sense in which the momentum of a billiard ball is independent of the paint on its surface. Mental properties are non-causally determined by neural properties.

If there is no law linking P to P^* , M may be partially causally responsible for P^* , and therefore exercise partial downward causal responsibility. In the absence of a law linking P directly to P^* , P^* 's occurrence cannot be lawfully explained by linking it to P . P and P^* are complex conjunctive neurophysiological properties. There may well be no law of neurophysiology, nor of chemistry or physics, that determines, as a function of P , the precise state P^* , expressed by a similar conjunction as P . The system might be too complex to show any regularity at that level of detail. The argument would of course be stronger if I could say "it is" rather than "it might". But first, there are empirical though certainly not conclusive reasons to think that the brain is a chaotic

²⁶ The distinction between M and P is neither an order-difference nor a level-difference in Kim's (1998) sense. M and P belong to the same level because they are exemplified by the same object; there is no difference of order because both are first-order properties.

²⁷ Marras (1998) argues that the mental/physical case is exactly analogous to the momentum/colour of the billiard balls case and that both relations of causal responsibility can exist in parallel without one interfering with the other, or being preponderant over the other.

system at the microscopic level²⁸. Second, the demonstration that it is a conceptual possibility suffices to refute Kim's argument, which aims at establishing the conceptual impossibility of the causal responsibility of mental properties.

Psychological laws

Our account depends crucially on the existence of genuine psychological laws. The best way to show that there are psychological laws, is to mention promising candidates. One such law bears on the cognitive procedures required for an animal to construct a representation of its environment, and to conceive its actions on the basis of this representation. Gallistel (1990) reviews ample evidence that otherwise extremely different animals, from the sand scorpion to man, navigate within their environment with the help of a "cognitive map". This map is built from egocentric representations (of the systems surroundings, as they appear to it from its present point of view), themselves constructed from sensory information, and bound together in a unique coherent geocentric metric map. One example of a lawful relation between mental properties M and M^* suggested by this work is the relation between the representation of the animal's own current position and direction of gaze (M) and the representation of the direction and distance to run to reach the nest (M^*). This computational relation is lawful (in the sense of a locally valid *ceteris paribus* law) and independent of the details of the neurophysiological network executing it.

Classical conditioning provides other phenomena subject to psychological laws. Rescorla and Wagner (1972) have established that the increase of associative strength between a conditioned stimulus (CS) X and an unconditional stimulus (US) in a trial in which X is presented just before US, is given by the formula: $\Delta V_X^n = \alpha_X \beta (\lambda - V_{AX}^{n-1})$,

where ΔV_X^n is the increase in associative strength of the CS X obtained in the n -th trial, α_X is the salience of stimulus X , β is the salience of the unconditioned stimulus, λ the maximal strength of association that can be obtained by association with the US²⁹, and V_{AX}^{n-1} is the total strength of association that has been reached between the US and two CS, X and A , in the preceding $n-1$ trials. This total strength V_{AX}^{n-1} is the sum of the strengths of the individual strengths of association of the two stimuli A and X , so that $V_{AX}^{n-1} = V_A^{n-1} + V_X^{n-1}$.

This formula, which expresses a law of learning, explains such phenomena of associative learning as masking and blocking: The learning of the association between A and the US can be "masked" by the presence of a second CS, X . Rescorla and Wagner's law predicts correctly that the associative strength of A grows more slowly in the presence of a second masking stimulus, X , than in situations in which the association between A and the US is learnt through presentations of A alone preceding the US; it also predicts correctly that the more salient is X the more the growth of A 's associative strength is slowed down. Another effect which can be shown to be a consequence of

²⁸ Cf. Skarda and Freeman (1990), Lehnertz and Elger (2000), Newman (2001).

²⁹ β is a measure of the speed of conditioning by association with the US, λ a measure of the asymptotic strength achievable in the long run.

Rescorla and Wagner's formula, is the fact that learning to associate A with the US is "blocked" by the existence of an (earlier established) association between another CS, X and the same US. The explanation is that $V_{AX}^{n-1} = V_X^{n-1}$ is already large when learning with A begins, and that the increase of associative strength ΔV_X^n is proportional to $\lambda - V_{AX}^{n-1}$, which is therefore small.

In this example, the earlier mental state M contains the associations between A , X and the US after the first $n-1$ trials and the experience of A preceding the US in the n th trial; the law predicts that this will produce a new mental state M^* containing associations whose strength is a function of M , as specified in the formula. This regularity is robust across animal species (Rescorla and Wagner report experiments with rabbits and rats), which makes it implausible that underlying a given state of learning, there is a unique microscopic property common to different test subjects belonging to different animal species³⁰.

To return to our question as to the relation between P , M , P^* and M^* : the law between M and M^* is certainly a "system law" whose validity is limited to living humans with the particular brain architecture characteristic of human beings³¹. In a given individual, M might have been determined by P_1, P_2, \dots instead of P . But given that M is actually determined by P , we may consider that part of the complex conjunction making up P constitutes an invariant constraint on the system whose stability makes it possible for the psychological law M - M^* to make the system evolve from M to M^* . The invariant part of P , which may be a general property of all humans or specific for the individual, remains stable throughout the evolution from M to M^* and determines which of many possible physical states P^*, P^*_1, P^*_2, \dots the system is in at the time at which it has M^* . This means that P^* is determined partly by P and partly by M although there is no downward causal law between M and P^* (as in Kim's option 3). Given that the initial state of the system is P , the law ruling the evolution of the system at the level of M constrains the system to evolve not only into an M^* state, but into an M^* state P^* compatible with the boundary conditions P .

It may turn out that many detailed states P^*_1, \dots, P^*_n are compatible with those constraints. This would mean that not even the combination of M and P is causally sufficient for P^* . In that case, *no* fact holding at the time of the P - M event would be causally responsible for the fact that the effect has precisely the micro-property P^* . All causally responsible facts of P^* would hold of events preceding P^* by a shorter time-span. Such a situation is typical for the evolution of complex systems. At the microscopic level, there is no fact causally responsible for the fact that the air molecules above Paris at time t move just as they do and why there are clouds exactly where they are, except facts about the same system a very short time before t . The complexity of the system is well over the threshold beyond which no lawful regularity allows long-term predictions at the microscopic level. We only look for causal explanations of facts about the large-scale patterns of weather because there are no facts bearing on much earlier times that are causally responsible for

³⁰ For other examples of realization-independent empirical psychological generalizations, Rey (1997), Antony and Levine (1997, p. 93/4) and Silverberg (2003).

³¹ On systems laws, Cf. Cartwright (1999) and Schurz (2002). The transition from M to M^* may be governed by an even more specific system law whose domain validity is only myself, maybe only during a certain period. It is an empirical question at which level there are lawful regularities.

the present detailed microscopic facts about the weather in Paris. In an analogous way, it is because there is no neurophysiological fact about my brain as it was yesterday or a year ago which is causally responsible for the brain state in which I now decide to close the window, that reference to mental states is necessary if we want to find causal explanations of mind-brain events at all.

This implies that causal responsibility is not transitive. There are causally responsible facts about the system's microstates at t for the system's microstates at $t + \Delta t$, but not for its microstates at much later times. Transitivity fails for lawful regularity, and indirectly for predictability and explicability. This consequence, far from being problematic, allows us to resolve a tension between conflicting intuitions about the transitivity of causation: most authors (cf. e.g. Lewis 1986) take the transitivity of causation simply for granted, but some (cf. e.g. Keil 2000) have noted that intuition is often reluctant to accept that an event is the cause of the indirect effects of its effects, especially if they are separated by a long chain of intermediary events. Both intuitions can be reconciled if causation is a transitive relation between events grounded on transference, whereas causal responsibility is a relation grounded both on transference and nomic dependency, which is not transitive because nomic dependency is not transitive.

Have we been led to deny the principle of the causal closure of the physical? Yes and no. There are cases of events (P^*-M^*) (such as bodily movements constitutive of actions) whose physical properties P^* are caused by the physical properties P of another event ($P-M$) only given that it also has mental properties M , where the M properties are themselves non-causally determined by properties among the physical properties P of the same event. Causation by mental events does not make mental properties any more spooky than other macroscopic properties: causally efficacious mental properties are properties which (as other macroscopic properties) are 1) non-causally determined by properties and relations at the physical level, 2) non-physical, in the sense of not being the subject matter of physical laws, but nevertheless 3) indispensable for causally accounting for certain physical events, such as bodily movements constitutive of actions. This account rejects the causal closure principle because according to 2) and 3), some events can only be causally accounted for by reference to non-physical, i.e. mental, properties. However, the account takes account of at least part of the intuitions lying behind the causal closure principle: those non-physical causes must themselves be physically determined, in the sense of being property instances that are non-causally determined by physical properties and relations.

Conclusion

My aim in this paper has been to sketch a conception of the relation of the mental properties of an individual to its underlying physical properties, which makes sense of the hypothesis that mental properties may be causally efficacious, while neither requiring them to be identical to physical properties nor to be irreducible with respect to them. The main claims are the following: physical properties of the parts of cognitive systems determine, by virtue of their nomological interactions, global properties of those systems, some of which are mental. These global properties

are first order-properties which may be causally efficacious by virtue of psychological laws. It is an empirical question whether their causal responsibility is paralleled on the microscopic level by neurophysiological laws, whose existence might threaten to make the mental properties epiphenomenal. It is conceivable that it is not so paralleled because the evolution of the brain is chaotic on a microscopic scale: it is plausible that there is no deterministic law governing its evolution to a precision comparable to that of the psychological laws, except for very short time spans. If the evolution of the cognitive system is governed only by psychological laws, the causal responsibility of its mental properties is vindicated. This is compatible with the reducibility of the mental properties by the discovery of laws of composition, which is a form of reduction that does not lead to identify mental properties to their neurophysiological base properties³².

References

- Achinstein, Peter (1974), The Identity of Properties, *American Phil. Quarterly* 11, p. 257-275.
- Antony, Louise M. and Levine, Joseph (1997), Reduction with Autonomy, in Tomberlin J.E. (ed.), *Philosophical Perspectives 11: Mind, Causation, and World*, Atascadero, CA: Ridgeview, p. 83-105.
- Baker, Lynne Rudder (1993), Metaphysics and Mental Causation, in: Heil and Mele (1993), p. 75-96.
- Baker, Lynne Rudder (1998), What We Do: a Nonreductive Account of Mental Causation, in Jan Bransen and Stefaan E. Cuypers (eds.), *Human Action, Deliberation, and Causation*, Dordrecht: Kluwer.
- Braddon-Mitchell, David and Jackson, Frank (1996), *The Philosophy of Mind and Cognition*, Blackwell, Oxford.
- Broad, C.D. (1925), *The Mind and its Place in Nature*, New York, Harcourt, Brace and Co.
- Campbell, Keith (1990), *Abstract Particulars*, Oxford, Blackwell.
- Cartwright, Nancy (1999), *The Dappled World, A Study of the Boundaries of Science*, Cambridge, Cambridge University Press.
- Chalmers, David J. (1996), *The Conscious Mind*. New York, Oxford University Press.
- Charles, David (1992), Supervenience, Composition, and Physicalism, in Charles and Lennon (1992), p. 265-296.
- Crane, Tim and D.H. Mellor (1990), There is No Question of Physicalism, *Mind* 99, p. 185-206.
- Crane, Tim (1995), The Mental Causation Debate, *Aristotelian Society, Suppl. Vol.* 69, p. 211-236.
- Cummins, Robert (1983), *The Nature of Psychological Explanation*, Cambridge (MA), MIT Press
- Dretske, Fred (1988), *Explaining Behavior*, Cambridge (MA), MIT Press, 1988

³² For critical comments on an earlier version of this paper, I am grateful to Alfredo Paternoster and Jürgen Schröder.

- Eells, Ellery (1991), *Probabilistic Causality*, Cambridge University Press, Cambridge.
- Ehring, Douglas (1997), *Causation and persistence: A Theory of Causation*, Oxford University Press, New York.
- Gallistel, Charles R. (1990), *The Organization of Learning*, Cambridge (MA), MIT Press.
- Glennan, Stuart S. (1996), Mechanism and the Nature of Causation, *Erkenntnis* 44, p. 49-71
- Hardcastle, Valerie Gray (1998), On the Matter of Minds and Mental Causation, *Phil. and Phen. Res.* 58, p. 1-25.
- Heil, John (1992), *The Nature of True Minds*, Cambridge, Cambridge University Press.
- Heil, John and Mele, Alfred (1993) (eds.), *Mental Causation*, Clarendon Press, Oxford.
- Horgan, Terence (1989), Mental Causation, in: Tomberlin J.E. (ed.), *Philosophical Perspectives 3: Philosophy of Mind and Action Theory*, Atascadero, CA: Ridgeview, 1989, p. 47-76.
- Jacob, Pierre (2002), Some Problems for Reductive Physicalism, *Phil. and Phen. Res.* 65, p. 647-653.
- Keil, Geert (2000), *Handeln und Verursachen*, Frankfurt a.M., Vittorio Klostermann.
- Kim, Jaegwon (1973), Causation, Nomic Subsumption and the Concept of Event, *J. of Phil.* 70, p. 217-236, in: Kim, J., *Supervenience and mind*, Cambridge, Cambridge University Press 1993, 3-21.
- Kim, Jaegwon (1978), Supervenience and Nomological Incommensurables, *American Phil. Quarterly* 15, p. 149-156.
- Kim, Jaegwon (1988a), Explanatory Realism, Causal Realism, and Explanatory Exclusion, *Midwest Studies in Philosophy* 12, p. 225-240.
- Kim, Jaegwon (1988b), Supervenience for multiple domains, *Philosophical Topics* 16, p. 129-150; repr. in Kim (1993c), chap. 7, p. 109-130.
- Kim, J. (1989) The myth of nonreductive materialism, in: Kim (1993c), p. 265-284.
- Kim, Jaegwon (1992), "Downward Causation" in Emergentism and Non-reductive Physicalism, in : Beckermann, Ansgar, H. Flohr, and J. Kim (eds.) (1992), *Emergence or Reduction ? - Essays on the Prospects of Nonreductive Physicalism*, Berlin, New York, de Gruyter, p. 119-138.
- Kim, Jaegwon (1993a), The Non-Reductivist's Troubles with Mental Causation, in: Heil and Mele (1993), p. 189-210, and in: Kim (1993c), chap. 17, p. 336-357.
- Kim, J. (1993b) Postscripts on Supervenience, in Kim (1993c), chap. 9, p. 161-171.
- Kim, J. (1993c), *Supervenience and mind*, Cambridge, Cambridge University Press.
- Kim, Jaegwon (1997), Does the Problem of Mental Causation Generalize ?, *Proceedings of the Aristotelian Society*, pp. 281-297.
- Kim, Jaegwon (1998), *Mind in a Physical World*. Cambridge (Mass.), MIT Press.
- Kistler, Max (1998), Reducing Causality to Transmission, *Erkenntnis* 48, p. 1-24.
- Kistler, Max (1999a), *Causalité et lois de la nature*, Paris, Vrin.
- Kistler, Max (1999b), Causes as events and facts, *Dialectica* 53, p. 25-46.
- Klee, Robert (1984), Micro-Determinism and Concepts of Emergence, *Phil. Sci.* 51, p. 44-63.

- Lehnertz, K. and C.E. Elger (eds.) (2000), *Chaos in Brain?*, World Scientific, Singapore.
- Lewis, David (1986), *Philosophical Papers, vol. II*, Oxford University Press.
- Lewis, David (1994), Reduction of Mind, in Guttenplan, Samuel (ed.), *A Companion to the Philosophy of Mind*, Blackwell, Oxford, p. 412-431.
- Loewer, Barry (2002), Comments on Jaegwon Kim's *Mind and the Physical World*, *Phil. and Phen. Res.* 65, p. 654-661.
- Lowe, E.J. (2000), *An Introduction to the Philosophy of Mind*, Cambridge, Cambridge University Press.
- Macdonald, Cynthia and Macdonald, Graham (1986), Mental Causes and Explanation of Action, in Stevenson, Leslie, Squires, Roger and Haldane, John (eds.), *Mind, Causation, and Action*, Blackwell, Oxford, p. 35-48..
- Machamer, Peter, Lindley Darden and Carl F. Craver (2000), Thinking about Mechanisms, *Philosophy of Science* 67, p. 1-25.
- Marras, Ausonio (1998), Kim's Principle of Explanatory Exclusion, *Australas. J. Phil.* 76, p. 439-451.
- McLaughlin, Brian (1993), On Davidson's Response to the Charge of Epiphenomenalism, in Heil and Mele (1993), p. 27-40.
- Newman, David V. (2001), Chaos, Emergence, and the Mind-Body Problem, *Australasian J. of Phil.* 79, p. 180-196.
- Putnam, Hilary (1992), *Renewing Philosophy*, Cambridge (MA), Harvard University Press.
- Rescorla, R.A. et Wagner A.R. (1972), A Theory of Pavlovian Conditioning : Variations in the Effectiveness of Reinforcement and Nonreinforcement, in : A.H. Black et W.F. Prokasy (eds.), *Classical Conditioning, II: Current Research and Theory*, p. 64-99, New York, Appleton-Croft-Century.
- Rey, Georges (1997), *Contemporary Philosophy of Mind*, Blackwell, Cambridge, MA.
- Robb, David (1997), The Properties of Mental Causation, *Phil. Quarterly* 47, p. 178-194.
- Schurz, Gerhard (2002), *Ceteris paribus* Laws : Classification and Destruction, *Erkenntnis* 57, p. 351-372.
- Silverberg, Arnold (2003), Psychological Laws, *Erkenntnis* 58, p. 275-302.
- Skarda, C. and W.J. Freeman (1990), Chaos and the New Science of the Brain, *Concepts in Neuroscience* 1, p. 275-285.
- Slors, Marc (1998), Two Claims that Can Save a Nonreductive Account of Mental Causation, in Jan Bransen and Stefaan E. Cuypers (eds.), *Human Action, Deliberation, and Causation*, Dordrecht: Kluwer.
- Sperry, R.W. (1976), Mental Phenomena as Causal Determinants in Brain Function, in: Gordon G. Globus, Grover Maxwell and Irwin Savodnik (eds.), *Consciousness and the Brain*. New York and London, Plenum Press, p. 163-177.
- Stephan, Achim (1992), The Historical Facets of Emergence, in: Beckermann, Ansgar, H. Flohr, and J. Kim (eds.) (1992), *Emergence or Reduction ? - Essays on the Prospects of Nonreductive Physicalism*, Berlin, New York, de Gruyter, p. 25-48.
- Thomasson, Amie (1998), A Nonreductivist Solution to Mental Causation, *Phil. Stud.* 89, p. 181-195.