MECHANISMS AND DOWNWARD CAUSATION

Max Kistler

Philosophical Psychology, vol. 22, No. 5, October 2009, 595-609.

Abstract

Experimental investigation of mechanisms seems to make use of causal relations that cut across levels of composition. In bottom-up experiments, one intervenes on parts of a mechanism to observe the whole; in top-down experiments, one intervenes on the whole mechanism to observe certain parts of it. It is controversial whether such experiments really make use of interlevel causation, and indeed whether the idea of causation across levels is even conceptually coherent. Craver and Bechtel have suggested that interlevel causal claims can be analysed in a causal and a non-causal component. I accept this idea but argue that their account should be modified so as to account of cases of apparent downward causation. First, constitution must be distinguished from identity; second, the analysis of downward causation requires the concept of a partial constraint. An analysis along these lines shows that the possibility of downward causation is not refuted by Kim's argument according to which it is incompatible with the completeness of physics.

1. Introduction

In the sciences and philosophy, questions often arise as to whether it is possible for phenomena at one level to interact causally with phenomena at another level. For some, it is obvious that contemporary science licences claims of causal interactions across levels. For others, the very idea of interlevel causation is conceptually confused. Recent work on the analysis of mechanistic explanation provides new conceptual tools for understanding how lower-level entities, such as components of a mechanism, may causally interact with higher-level entities, such as phenomena at the level of the mechanism as a whole. I will build on a proposal by Craver and Bechtel (2007) who analyze interlevel causal claims as hybrid, part causal and part constitutive. I accept their idea that interlevel causal claims can be analysed in a causal and a non-causal part, but disagree about their interpretation of the non-causal part. I will argue first that levels are related by constitution, not identity, and second, that constitution is not the appropriate concept to account for downward causation. Third, I propose an analysis of downward causation according to which higher level phenomena constrain phenomena at lower levels. Finally, I argue that this account avoids well-known

objections such as Kim's who has argued that downward causation conflicts with the metaphysical principles of the closure of the physical domain and of causal exclusion.

According to the mechanistic conception of the scientific explanation of the behaviour of complex systems, such an explanation typically takes the form of showing how the parts of a system, thanks to their articulation and interactions, give the whole system certain properties. A mechanism is "a set of entities and activities organized such that they exhibit the phenomenon to be explained" (Craver 2007, p. 5). The aim is often to explain processes: dynamical mechanisms are "productive of regular changes from start or set-up finish or termination conditions" (Machamer, Darden, and Craver, 2000, p. 3).

Mechanistic analyses typically make reference to several levels of organisation. The top level corresponds to the mechanism as a whole, as well as to its properties and activities. One cannot e.g. understand the mechanism of long-term memory without speaking of what it achieves at the level of the behaviour of an animal: it modifies its behavioural capacities[1]. Consider a mouse learning to navigate in a Morris water maze[2]. This learning is mediated by the modification of the synaptic structure of the hippocampus which is thought to develop a "spatial map" of the maze. The hippocampus is an essential part of the mechanism, and situated at a second level, relative to a first and highest level, which is taken to be the whole organism. The modification of single synapses is a crucial part of the mechanism leading to modification of the synaptic structure of the hippocampus. Such synapses, being essential parts of the hippocampus required for its function within the mechanism of long-term memory, lie at a third level. Finally, the modification of a synapse requires modification at a still lower fourth level, which corresponds to molecular parts of the synapses, in particular specific receptor molecules that control the properties of the neurons' cellular membranes. Here, levels are defined locally with respect to a given mechanism and its component parts. Things and activities are at different levels if and only if they "stand in part-whole relations to one another with the important additional restriction that the lower-level entities and activities are components of the higher-level mechanism" (Craver and Darden 2001, p. 117). A part is a component if it makes an identifiable contribution to the activities of the whole mechanism[3].

---

1 Here I follow the analysis of the mechanism of long-term memory by Craver and Darden (2001).

2 Since its introduction in 1984, the Morris water maze has become a major experimental tool for the behavioural investigation of spatial learning and memory. Rats or mice are placed in a circular pool filled with opaque water, which hides from their view a small platform under the surface of the water, on which they can stand. Rats swim well but prefer to stay on the platform. Experiments study the rats' learning of the location of the platform in different circumstances.

3 This notion of mechanistic levels is developed in Craver (2007, chapter 5).

2. "Top-down" and "bottom-up" experiments

The experimental investigation of mechanisms can quite naturally be interpreted in terms of causal processes crossing the boundary between levels of composition, both in upward and downward direction[4].

In "bottom-up" experiments, one manipulates properties ("independent variables") of individual components of a mechanism in order to observe the consequences of this intervention at the level of system properties ("dependent variables"), i.e. properties belonging to the whole mechanism. An important category of bottom-up experiments uses the so-called "knockout" technique: organisms are genetically modified in such a way that specific genes are deleted. The observation of the development and behavioural capacities of such animals is taken to license inferences about the causal contribution of the knocked out genes to the development and capacities of the animal. It has, e.g., been experimentally discovered that mice in which the gene that codes for the NMDAR1-subunit of the NMDA-receptor[5] is knocked out selectively in the relevant region of the hippocampus[6], exhibit characteristic deficits in the acquisition of long-term memory: They have, e.g., difficulties learning the location of the submerged platform in a Morris water maze. In such experiments, a causal intervention at the molecular level seems to have effects at the level of the learning capacities of the organism. It seems to be a case of "bottom-up causation".

In downward or "top-down" experiments, the experimental intervention consists in manipulating system properties and observing the effects of this intervention on properties of components of the mechanism. Such experiments seem to presuppose the possibility of top-down causation. In techniques of brain-imaging such as fMRI (functional magnetic resonance imaging) and single-cell recording, the experimenter manipulates system properties, e.g. by putting animals in a situation in which they accomplish a specific behavioural task, and observes subsequent modifications of properties at lower levels: fRMI allows to measure nervous activity in specific brain regions; single cell recording allows to observe the activity of individual neurons[7]. It has, e.g., been discovered that the CA3 and CA1 regions of the hippocampus of rats and mice contain so-called place cells that fire intensely only when the animal's head is in a certain part of the environment called the cell's "place field"[8]. Such

---

4 Craver (2002) analyses such interlevel experiments within a mechanistic approach.
5 NMDA is short for N-Methyl D-Aspartate.
6 Nakazawa et al. (2004).
7 Ludvig et al. (2001).
8 Muller (1996).

experiments causally intervene at the level of the organism: one manipulates the behaviour of the whole animal. The measured effect of that intervention lies at the level of the animal's microscopic constituents: One observes modifications of the properties and activities of neurons in the hippocampus.

Are such "interlevel" experiments instances of interlevel causal relations? Scientists' statements suggest an affirmative answer. In Eric Kandel's words, the "biological analysis of learning requires the establishment of a causal relation between specific molecules and learning" (Kandel, 2000, p. 1268). More specifically, Kandel takes bottom-up causation to be required to understand how changes in the brain can affect cognitive processes and behaviour. In particular, "certain diseases ('organic diseases') affect mentation through biological changes in the brain" (*ibid.*). Kandel also acknowledges the existence of downward causation: "Learning produces changes in the effectiveness of neural connections" (Kandel, 2000, p. 1275). Downward causation also seems to be required to make sense of psychotherapy: "Insofar as social intervention works […] [e.g.] through psychotherapy […] it must work by acting on the brain" (*ibid.*).

Recent philosophical work on causation also seems to lead to acknowledge bottom-up and top-down causation. According to Woodward (2003), causation can be analysed in terms of manipulability. If a cause of some event, property or factor *e* is a factor *c* such that interventions on *c* allow manipulating *e*, then the bottom-up and top-down manipulations undertaken to understand the working of mechanisms are cases of causation.

However, downward causation, through which the state or activity of a complex system causally influences its own parts, is controversial. Craver and Bechtel's reason for thinking that "top-down causation is incoherent" (Craver and Bechtel 2007, p. 547) is that they interpret downward causation as a wrongheaded attempt to interpret the relation between a whole mechanism and one of its components, where both are *taken at the same moment*. A relation between a state of the whole and the state of one of its components at the same moment can indeed not be causal. It is part of the concept of causation that causes must be spatially and temporally distinct, so that they cannot have any common spatial or temporal parts, and also that causes must precede their effects. It would indeed "violate many of the central ideas associated with the concept of causation" (Craver and Bechtel 2007, p. 554) to consider as causal a *simultaneous* relation between a state of a whole and the state of one of its parts. The question I address here is whether there can be "delayed" top-down causation: Can a higher-level event, such as a mental event *C* happening in a person at time t_1, influence

an event that happens at some later time t_2 and at a lower-level, i.e. at the level of certain components of the person?

Jaegwon Kim (1998) has argued that the hypothesis of downward causation in this sense leads to insuperable conceptual difficulties and can thus be refuted by reductio. According to Kim, downward causation is conceptually incompatible with two metaphysical principles: the principle of causal closure of the physical domain ("Closure", for short) and the principle of explanatory exclusion ("Exclusion", for short). Given Closure, if a physical event *e* at t_2 has a cause at some earlier time t_1, then it has a physical cause *c* at t_1. But then, given Exclusion, the existence of a physical cause *c* is incompatible with the existence of any higher-level, non-physical cause *C* at t_1. I shall examine Kim's argument later. It puts us before a dilemma: Either the argument is sound, and we must revise our interpretation of interlevel experiments on complex systems so that it does not require any downward causation after all, or we abandon one of the two metaphysical principles Kim uses, so as to open up the logical space for downward causation.

3. Reducing causation to mechanism?

No doubt, many mechanistic explanations are causal explanations, insofar as they explain the activity of a mechanism that brings about, or causes, a termination condition from some initial condition. It is clear that initial and termination conditions are meant to bear on different moments in time. Hence there can be no question of a "mechanism" linking two aspects of the same event. As a consequence, a mechanistic analysis avoids the wrong result of the traditional deductive-nomological account of causation[9] that there may be causal relations between different properties of one substance at one time, such as between the temperature and the pressure of a given sample of gas. However, it would be too quick to conclude that there cannot be causation across levels for the reason that there cannot be a mechanism mediating between the activity of a whole and an entity or activity that is a part of that whole. It would be too quick because not all causation is mechanical causation. It is not true that causation can be reduced to mechanism, as Stuart Glennan suggests when he writes: "Events are causally related when there is a mechanism that connects them" (Glennan, 1996,

[9] According to this account, the statement that event *a* causes event *b* means that there are laws of nature, which make it possible to deduce a description of *b* from a description of *a*. "What is meant when it is said that event *b* is caused by event *a*? It is that there are certain laws in nature from which event *b* can be logically deduced when they are combined with the full description of event *a*." (Carnap 1966, p. 194; variables modified). For a criticism, see Kistler (2006, chap. 5).

p. 49)[10]. Causation cannot be reduced to the relation between the initial and end condition of a mechanism because there are causal relations that are not mediated by any mechanism. Glennan's mechanical account "cannot explain causation in fundamental physics" (1996, p. 50). It cannot be true of interactions between elementary particles that the existence of a causal relation is equivalent to the existence of a mechanism. Glennan concludes that there are two fundamentally different kinds of causation and suggests that "there should be a dichotomy in our understanding of causation between the case of fundamental physics and that of other sciences." (1996, p. 50)

However, one would need stronger reasons to justify the radical and counterintuitive conclusion that there are two distinct concepts of causation, one for fundamental physical interactions and one for all other causal relations. This consequence is avoided as soon as one abandons the idea that causation can be *reduced* to mechanism. On closer inspection, it appears that the concept of mechanism presupposes that of causation, far from being reducible to it. Providing a mechanistic explanation means to decompose the working of a complex system into a number of simpler subsystems that interact causally with each other. These subsystems can in general themselves be analysed in still simpler subsystems, so that the interactions between the former subsystems can also be mechanistically explained. The crucial point is that each step of the analysis of a mechanism makes essential use of the notion of cause, and thus presupposes it. If one pushes the analysis far enough, one eventually reaches interactions between elementary particles. These however cannot in their turn be given a mechanistic analysis, because elementary particles cannot be decomposed into their parts. It follows that the concept of mechanism cannot be used to analyse the concept of causation and that, quite on the contrary, the concept of causation is among the irreducible conceptual instruments of mechanistic analysis. Mechanist causation rests in the last instance on the causation of fundamental physical processes[11].

4. Analysing interlevel causation in terms of constitution

Craver and Bechtel (2007) have put forward an analysis of interlevel causal claims, according to which they are composed of an intralevel causal part and a non-causal interlevel

[10] This is meant to be a metaphysical claim, which is very different from Ahn and Kalish's (2000) psychological claim that subjects intuitively ground their causal judgments on information on mechanisms.

[11] Thus, as Craver (2007, p. 90/1) notes, there is a tension between Glennan's claims that "causal talk about interactions governed by fundamental laws" is "problematic" (Glennan, 1996, p. 67), and that causation can be reduced to mechanism: The causal relation between the initial and end states of a mechanism is ultimately grounded on the causality of the fundamental interactions.

part. Thus, they answer Kim's challenge by analyzing interlevel causation away. The interlevel constituent of apparent interlevel causal claims is really non-causal. It is constitution. After all, Craver and Bechtel argue, one can make sense of interlevel experiments on complex systems without positing downward causation in the strict sense. When a part of a mechanism is manipulated by intervening on a system property, the relevant downward relationship, they argue, "is not a *causal* relationship" (2007, p. 547; their emphasis). In what follows, I shall first present Craver and Bechtel's proposal to "analyse away" apparent cases of upward and downward causal relations and argue that their analysis is not adequate in all cases. If I am right, then in order to take account of some recalcitrant cases of downward causation, it is inevitable to choose the second horn of the dilemma mentioned above, and resist Kim's verdict against downward causation. Second, I shall sketch a general reason to abandon Kim's principle of the causal closure of the domain of physical phenomena, which is also a reason to acknowledge the conceptual possibility of downward causation.

Craver and Bechtel (2007) argue that all apparent instances of bottom-up or top-down causation can be conceived of as composed of intralevel causation and non-causal determination. However, their interpretation of this determination relation in terms of constitution is flawed in the case of downward causation: top-down determination cannot be analyzed in terms of constitution.

Let me begin by the metaphysical notion of constitution. In the present context, constitution is a relation between a macroscopic object and a complete collection of its parts. It applies to mechanisms as a special case: a mechanism is constituted by its components at a given level of decomposition. However, as the constitution relation applies not only to mechanisms, I will use Unger's (1980) example of the relation between a cloud and the droplets it contains. The same points could be made with any other macroscopic object, whether it can be mechanistically analysed or not. One might be tempted to say that a given cloud is identical with the collection of all the tiny droplets it contains because, at a given time, the cloud and the collection of drops share all their parts. However, here are two reasons to think that they are not identical[12]: first, considering the evolution of the cloud in time, the concept of cloud allows it to persist, i.e. to continue to exist and remain the same cloud, while individual drops enter or leave it. However, each time a drop is added or removed, the collection of drops in the cloud changes. Moreover, and this is the second reason for

[12] It is controversial whether these reasons are sufficient to deny that constitution is identity. Cf. Rea (1997).

distinguishing the cloud from the collection of its drops, even at a given moment of time, it would have been *possible* that the very same cloud contains some more drops or some less. Let us admit Kripke's thesis that all true identity statements of form "A=B", where A and B are rigid designators, are *necessarily* true. The contrapositive of this thesis is that if a statement attributes a contingent relation to A and B, that relation is not identity. The fact that there could have been a different collection of drops in the cloud shows that the relation between the collection of drops and the cloud is contingent. Therefore it cannot be a relation of identity. Here is where constitution steps in: One can say that the actual collection of drops constitutes the cloud although they are not identical.

Three features of constitution will prove important in what follows. First, it is asymmetric[13]: if A constitutes B, it is impossible that B constitutes A. The collection of drops constitutes the cloud but the cloud doesn't constitute the collection of drops. Second, a given object can be, successively or alternatively, constituted by more than one collection of parts. One might express this by saying that some objects allow for "multiple constitution". Third, constitution is a relation of logical and metaphysical, rather than epistemic or nomological type.

Let us now turn to Craver and Bechtel's analysis of apparent cases of downward causation. Take their example of the process that begins with a person's decision to start a tennis game and leads to appropriate tennis-playing behaviour. The latter requires increased glucose consumption in the person's muscle cells. The decision, a system property of the person, seems to have effects at the cellular and molecular levels. However, Craver and Bechtel argue that this appearance is misleading, and dissolves at closer inspection. "The case can be described without remainder by appeal only to intralevel causes and to constitutive relations" (Craver and Bechtel, 2007, p. 559). If this is correct, downward causation can be analysed according to one of two patterns.

FIGURE 1 HERE

[13] This makes Craver and Bechtel's claim puzzling that the constitution relation between a mechanism and one of its components is "symmetrical". Their justification seems to show that they take constitution to be equivalent to identity, which is indeed symmetrical: "A change in the parts is manifest as a change in the mechanism as a whole, and a change in the mechanism is also a change in at least some of its component parts" (Craver and Bechtel 2007, p. 554).

In scenario 1, *C* (the decision) determines *c* (the brain state underlying the decision), which then causes *e* (enhanced consumption of glucose in muscle cells) by intralevel causation.

FIGURE 2 HERE

In scenario 2, *C* (as before) causes *E* (appropriate behaviour at the level of the organism), which then determines *e* (as before) in a non-causal way.

The first scenario is inadequate if, as is generally assumed, mental events such as decisions to play tennis are multirealisable by many different brain states. Which particular brain state *c* realises *C* depends on the person's history and the circumstances. At any rate, *C* does not by itself determine *c*. Furthermore, even if it did (in other words, if we abstract away from multiple realisation), the downward determination of a brain event by a mental event could not possibly be construed as a relation of constitution, because constitution is a bottom-up relation.

The same reasons seem to make scenario 2 inadequate: First, *E* does not in itself determine *e* because tennis-playing behaviour, and even a given detailed bodily movement, can be realised at the molecular level in many ways. Second, *E* does not constitute *e*: Parts can be constitutive of wholes but wholes cannot be constitutive of their parts.

5. Downward causation and downward constraints

However, it is possible to reinterpret scenario 2 in such a way that it may represent the situation correctly. This requires modifying Craver and Bechtel's proposal in two respects. First, the downward relation by which *E* determines *e* is a relation of *constraint* not of constitution. Second, the constraint imposed on *e* by *E* is not complete but partial. Here is what makes the notion of constraint useful for evaluating the conceivability of downward causation. A constraint limits the possibilities of evolution or change accessible to a system. In a system of equations with *n* variables, each equation imposes a constraint on the variables, in the sense of limiting the values the variables can take to satisfy the equations. If the variables represent the degrees of freedom of a physical system, i.e. the dimensions within which the state of the system can evolve, the notion of constraint acquires a physical meaning. Each equation expressing a link between the variables expresses a limitation imposed on the

possibilities of evolution of the system[14]. A ball rolling down an inclined plane is constrained to rest on that plane: The possibilities of its movement are limited to the two dimensions of the plane. In a similar manner, properties of complex systems can be sources of constraints imposed on its constituents. By maintaining the temperature of a gas at a fixed value, one exercises a constraint on the mean kinetic energy of the molecules constituting the gas. The impermeability of its container imposes a constraint on the number of molecules. Each constraint on a macroscopic system diminishes the number of possible states of its constituents. However, as long as there are fewer constraints than degrees of freedom, the constraints on a system determine its state, and the state of its parts, only partially.

Contrary to constitution, constraint is not an asymmetric relation. One can say that the state of the parts of a system constrains the state of the whole[15]; but it can also be correct to say that the state of the whole constrains the states of the parts, as when the temperature or the volume of the container of a gas limits the degrees of freedom of the molecules constituting it.

The notion of degree of freedom, and thus the notion of a constraint limiting the degrees of freedom of a system, can be generalized to all determinable properties of a system that can take different values. An animal's body temperature corresponds to a degree of freedom subject to the constraint of remaining within limits imposed by a regulatory mechanism at the level of the organism. However, this temperature itself imposes a constraint on the possible states of motion of the molecules composing the organism. The overall temperature imposed on the body by the regulatory mechanism limits the space of possible states of motion of the body's constitutive molecules, by fixing the mean kinetic energy of their states of motion. In the same sense, the fact that a given cognitive system is at a given moment in some cognitive state, e.g. of consciously perceiving an approaching tennis ball, imposes a constraint on the possible states of its parts, and first of all on the state of its neurons. It is incompatible with many neuronal states, such as states corresponding to closed eyes or the contemplation of an immobile scene. However, the constraint is only partial because it is compatible with many microscopic states of neurons and molecules.

The process leading from the decision (*C*) of a person to her playing tennis (*E*) is an intralevel causal process at the level of the organism. I suggest that the concept of *partial constraint* helps us understand the relation between tennis playing and the underlying

---

[14] In classical mechanics, Lagrange has provided an algorithm for finding the trajectory of a system that is subject to constraints, on the basis of the representation of each constraint by an equation.

[15] Shapiro (2005) shows how the structure of the brain, shaped by phylogenetic and ontogenetic evolution, contrains cognitive functions, and thus sets limits on the multirealizability of those functions. These are bottom-up constraints.

microscopic events *e* taking place in the body, such as enhanced glucose uptake in muscle cells. The state of organism *E* exerts a constraint on its parts, in the sense that the fact that the organism is in state *E* limits the space of possible states of its muscle cells. However, the detailed evolution of each muscle cell is also constrained at the cellular and molecular level, by the physical state of the cell and its surrounding.

My suggestion to conceive of the influence by *E* on *e* as a partial constraint seems to resemble what Craver and Bechtel call the "enlisting" of the system's parts by the whole system. However, they interpret enlisting to be a form of identity or constitution. To say that the tennis player's "glucoregulatory mechanisms are *enlisted*" (Craver and Bechtel, 2007, p. 559; their emphasis) by his activity, is according to them equivalent to saying that "a change in the activity of the mechanism as a whole just is a change in one or more components" (*ibid.*). However, the fact that many detailed states of the parts are compatible with a given state of the whole makes identity ("just is") inappropriate to characterise the relation between *E* ("a change in the activity of the mechanism as a whole") and *e* ("a change in one or more components"). The constraint exercised by *E* on *e* is a form of *partial* determination: It is only together with other constraints that lie at the cellular and molecular levels in the particular situation that *E* determines one particular microscopic change *e*.

Craver and Bechtel say that enlisting is a form of constitution: the activity of playing tennis "is in part constituted by activities at neuromuscular junctions" (Craver and Bechtel, 2007, p. 559). If my argument above is correct that constitution is not identity, one cannot analyse enlisting both as constitution and as identity. However, even if we ignore that controversial issue here, the concept of constitution is inadequate for the relation between a state of the system, such as performing the action of hitting a tennis ball, and the collection of states of the parts of the system, for two reasons. First, the parts constitute the whole but the whole cannot be said to constitute its parts. Second, the (partial) determination of the state of the parts by the state of the whole is not conceptual but rests on laws of nature. It is clear that empirical research is necessary to find out which states of its neurons and their connections are compatible with, e.g. the fact that the system consciously perceives an approaching tennis ball.

The notions of constitution and constraint, which are both forms of non-causal determination, make causal relations crossing levels of composition conceivable. It is after all conceptually possible that a change occurring at the level of the parts of a system at time t_1 causes changes at the level of its systemic properties at some later time t_2, and it is also possible that a change of systemic properties at time t_1 causally influences the states of its

parts at some later time t_2. Just as Craver and Bechtel, I have tried to analyse downward causation in terms of a causal and a non-causal part. There are at least two reasons for thinking that constraints are a form of non-causal determination. If the macroscopic state of a system exercises a constraint on the state of one of its parts, the origin of the constraint (the state of the system) and what it constrains (the state of the part) occupy the same place at the same time. But first, cause and effect must not overlap in space, and second, causes must precede their effects. Therefore, constraints do not cause what they constrain. However, these are not sufficient reasons "to be sceptical of interlevel causal claims" (Craver 2007, p. 182). Top-down causation is conceivable if it is understood in terms of the influence of a higher level property *C* of a whole mechanism at time t_1 on a property *e* of a lower-level part at some later time t_2. If a higher-level law imposes a constraint on the state of system S, a higher-level property of S at time t_1 can be (partly) causally responsible for a lower-level property that one of the components of S has at some later time t_2.

With this analysis in mind, let us get back to Kim's argument against the possibility of downward causation[16]. According to Kim, the idea that a change in system properties might exercise a causal influence on the properties of the system's parts is incompatible with two metaphysical principles. The principle of the causal closure ("Closure") of the domain of physical phenomena says that for a given physical event *e* that takes place at time t_2, for each time t_1 preceding t_2, there is a complete physical cause *c* (at t_1) of *e*[17]. The principle of explanatory exclusion ("Exclusion") says that there is no systematic overdetermination of microscopic events by independent micro- and macroscopic events. If event *e* at t_2 has a complete physical cause *c* at time t_1, then it does not (at least not in the general case) in addition have another complete cause *C* at the same time t_1, which is independent of *c*.

Kim's argument establishes the impossibility of downward causation of lower-level effect *e* by higher-level cause *C* in two steps. In the first step, Kim argues that, given that the state *e* at time t_2 is a physical event, Closure guarantees that *e* has a complete physical cause, *c*, for each time t_1 preceding t_2 at which it has a cause at all[18],.In the second step, he argues that, given Exclusion, *e* cannot have any causal explanation independent of its microscopic causal explanation in terms of *c*; this excludes in particular the possibility of a causal explanation of *e* in terms of the mental cause *C*.

---

16 I have analysed Kim's argument in more detail in Kistler (2005) and (2006a).

17 Cf. Kim (1998, p. 37/8). See also Lowe (2000; 2000a, p. 26ff.).

18 The last clause is supposed to allow for the possibility that there may be times t at which e has no cause at all, in the sense that there are not, at t, any conditions that are nomologically sufficient for e.

The principle of the causal closure of the physical domain can be doubted. Downward causation is possible if there can be microscopic events in complex systems that are not completely determined, in the long run, by same-level events. Cellular or molecular changes in a living organism may, e.g., not be completely determined over long time intervals by other cellular or molecular events. In that case, one cannot (deductively) explain a molecular event in a living organism (such as the hydrolysis of an ATP molecule in order to deliver the energy necessary for muscle contraction), on the basis of other molecular events that have occurred much earlier. It cannot of course be proven that such a deductive explanation is impossible in principle. But it cannot at present be done; rather, the scientific explanation of such a microscopic fact in terms of the state in which the system has been in much earlier, will mention processes at higher levels. In the present state of scientific knowledge, any explanation of the molecular change in terms of causes preceding the event by more than a few milliseconds must mention higher-level mechanisms in which that change is a component. The person of which these molecules are components engages in some purposeful physical activity, such as playing tennis; and this activity has a psychological explanation. The muscular effort required for such a physical activity requires energy, and the hydrolysis of ATP delivers energy. My claim is only that the burden of proof is on those who claim that it is "in principle" possible to explain the hydrolysis of the ATP molecule in terms of causes much earlier, directly at the chemical and physical level, without any reference to higher level properties. Present day science is unable to do this; furthermore, scientists seem to consider that this unability is not worth the effort to be overcome. So what might be the ground for claiming it might in principle be overcome?

The same reasoning can help justifying how a metaphysical conclusion – that the state *e* of the collection of parts of the system at t_2 is not directly causally determined, at this level of parts, by the state of the collection of parts of system at t_1 – can follow from an epistemic premise – that it is impossible to explain or predict *e* at t_2 (fact B) from knowledge of state *c* at time t_1 (fact A), together with low-level physical laws. One cannot of course in general conclude from the absence of a causal explanation of fact B by fact A that A does not causally determine B. Maybe we just ignore the objective determination of B by A. However, in some cases, we have scientific reasons to think that there is objectively no causal determination between successive states at a given level. Or rather, we have no scientific reasons to believe that there is such a causal determination. In systems whose evolution is described by quantum mechanics, the outcomes of some measurements at t_2 are not determined by the system's state at t_1. Deterministic chaos can also make the causal explanation of the state of the system at t_2

by its state at t_1 in principle impossible. No empirical sense can be attached to the hypothesis that a determinable property of a physical system with a continuous value pattern, possesses at time t_2 an absolutely precise value. There are absolute limits to the possible precision of measures that appear in the so-called uncertainty relations of quantum mechanics. Even if the state of a chaotic system has been determined with the absolutely maximal precision at time t_1, that state does not completely determine the state of the system at times t_2 that are sufficiently distant from t_1. In such a system, the "horizontal" determination of physical events at the physical level is objectively incomplete. This throws doubt on Closure. However, the state of a system whose evolution has no complete physical determination may well be completely determined nevertheless. The success of ethology and psychology in explaining numerous animal and human behaviours shows that animals and humans obey "system laws"[19] constraining their evolution at the level of systemic properties, such as cognitive laws determining actions on the basis of reasoning and decision-making. The fact that an organism obeys such laws means that its evolution obeys constraints at a psychological level. The constraints exercised on the organism by laws at different levels, at the level of the organism as a whole, and at various lower levels corresponding to its parts, create no conflict. If the determination of a molecular event is incomplete at its own level, it may nevertheless be completely determined jointly by laws at molecular and system levels. A given molecular event happening in an organism may be partly determined by constraints at the molecular level and partly by downward constraints from the psychological level, insofar as the organism obeys psychological laws.

FIGURE 3 HERE

This scenario[20], sketched in fig. 3, expresses the thesis that, contrary to Kim's closure principle, present-day scientific knowledge does not exclude the possibility that the domain of physical phenomena is not closed. The microphysical state of a complex system at t_1 may not completely determine its microphysical state at a much later time t_2. In such a system, the microphysical state at t_2 may be partially determined in a downward direction by the constraint that the system must, at t_2, be in a global state compatible with system level laws,

[19] Cognitive laws linking actions to reasoning and decision are one case of what Cummins (2000) calls "laws *in situ*" and what Schurz (2002) calls "system laws". Insofar as an organism exhibits regularities at the level of the organism, it is what Cartwright (1999) calls a "nomological machine".

[20] I have justified this sketch in a little more detail in Kistler (2006a).

such as cognitive laws. The determination of state *e* is completed by the physical circumstances *d* occurring immediately before *e*.

6. Conclusion

The experimental investigation of mechanisms seems to presuppose the existence of causal relations crossing levels of composition. Bottom-up and top-down experiments suggest that there can be both upward causation where parts influence a whole mechanism, and downward causation where the state of a mechanism influences the state of its parts. Many philosophers take such "downward causation" to be incompatible with general metaphysical principles abstracted away from science, such as the principle of the causal closure of the domain of physical events. I have tried to show that partial downward determination of the properties of parts of mechanisms by properties of the whole mechanism is conceivable and does not violate any plausible scientific or metaphysical principles.

My proposal for understanding the possibility of downward causation builds upon Craver and Bechtel's (2007) claim that apparent cases of downward causation can always be analysed in terms of intralevel causal relations and relations of constitution. I have suggested modifying their analysis in two ways. First, constitution is not identity: The possibility that the parts of a system change whereas the systems remains the same shows that the collection of parts constituting a whole is not identical to the whole. Second, the analysis of downward causal judgments requires the notion of constraint rather that that of constitution. Furthermore, I have argued that the judgment that a given lower-level property *e* of a part of a mechanism is causally influenced in a "downward" way by a higher-level property *C* of the whole mechanism, can be empirically justified if the system obeys to a system law at the level of *C*, whereas there are no empirical reasons to believe that the evolution of the part is completely determined at its own level[21].

References

Ahn, W.K. & Kalish, C.W., (2000). The Role of Mechanism Beliefs in Causal Reasoning, in F. Keil & R. Wilson (eds.), *Explanation and Cognition*, Cambridge, MA: MIT Press, p. 199-225.

[21] I thank Reinaldo Bernal and Carl Craver, as well as my auditors at Geneva and Madrid, for helpful criticism of earlier versions of this paper.

Carnap, Rudolf (1966), *Philosophical Foundations of Physics. An Introduction to the Philosophy of Science*, Martin Gardner (ed.), New York, Basic Books.
Cartwright, N. (1999). *The Dappled World: A Study of the Boundaries of Science*. Cambridge: Cambridge University Press.
Craver, C.F. (2002). Interlevel Experiments and Multilevel Mechanisms in the Neuroscience of Memory. *Philosophy of Science*, 69, S 83-S 97.
Craver, C.F. (2007). *Explaining the Brain: Mechanisms and the Mosaic Unity of Neuroscience*. Oxford: Oxford University Press.
Craver, C.F. & Bechtel, W. (2007). Top-down Causation Without Top-down Causes. *Biology and Philosophy*, 22, 547-563.
Craver, C.F. & Darden L. (2001), Discovering Mechanisms in Neurobiology: The Case of Spatial Memory, in P.K. Machamer, R. Grush, and P. McLaughlin (eds.), Theory and Method in the Neurosciences, Pittsburg, PA: University of Pittsburgh Press, p. 112-137.
Cummins, R. (2000), How Does it Work? Vs. What Are the Laws? Two Conceptions of Psychological Explanation, in F. Keil & R. Wilson (eds.), *Explanation and Cognition*, Cambridge, MA: MIT Press, p. 117-45.
Glennan, S. (1996). Mechanisms and the Nature of Causation. *Erkenntnis*, 44, 49-71.
Kandel, E.R (2000). Cellular Mechanisms of Learning and the Biological Basis of Individuality. In E.R. Kandel, J.H. Schwartz, & T.M. Jessell (Eds.) *Principles of Neural Science* (pp. 1247-1279). New York: McGraw-Hill.
Kim, J. (1998). *Mind in a Physical World*. Cambridge, MA: MIT Press.
Kistler, M. (1998). Reducing Causality to Transmission. *Erkenntnis*, 48, 1-24.
Kistler, M. (2005). Is Functional Reduction Logical Reduction?. *Croatian Journal of Philosophy*, 5, 219-234.
Kistler, M. (2006). *Causation and Laws of Nature*. London: Routledge.
Kistler, M. (2006a). The Mental, the Macroscopic, and their Effects. *Epistemologia*, 29, 79-102.
Lowe, E.J. (2000). Causal Closure Principles and Emergentism. *Philosophy*, 75, 571-585.
Lowe, E.J. (2000a). *An Introduction to the Philosophy of Mind*. Cambridge: Cambridge University Press.
Ludvig, N., Botero, J.M., Tang, H.M., Gohil, B. & Kral, J.G. (2001). Single-cell recording from the brain of free moving monkeys. *Journal of Neuroscience Methods* 106 (2), 179-187.
Muller, R. (1996). A Quarter of a Century of Place Cells. *Neuron*, 17, 813-822.
Nakazawa, K., McHugh, T.J., Wilson, M.A. & Tonegawa, S. (2004). NMDA Receptors, Place Cells and Hippocampal Spatial Memory. *Nature Reviews Neuroscience*, 5, 361-372.
Rea, M. (ed.) (1997). *Material Constitution*. Lanham, MD: Rowman and Littlefield.
Schurz, G. (2002). *Ceteris paribus* Laws: Classification and Deconstruction. *Erkenntnis*, 57, 351-372.
Shapiro, L.A. (2005). *The Mind Incarnate*. Cambridge, MA: MIT Press.
Unger, P. (1980). The Problem of the Many. In P. A. French, T. E. Uehling, Jr., & H.K. Wettstein (Eds.) *Midwest Studies in Philosophy*, 5 (pp. 411-67). Minneapolis: University of Minnesota Press.
Woodward, J. (2003). *Making Things Happen*. Oxford: Oxford University Press.