Horizontal and Vertical Determination of Mental and Neural States

Jens Harbecke
Witten/Herdecke University
Alfred-Herrhausen-Str. 50, D–58448 Witten

Harald Atmanspacher
Institute for Frontier Areas of Psychology
Wilhelmstr. 3a, D–79098 Freiburg

Abstract

Mental and neural states are related to one another by vertical (synchronic) interlevel relations and by horizontal (diachronic) intralevel relations. For particular choices of such relations, problems arise if causal efficacy is ascribed to mental states. In a series of influential papers and books, Kim has presented his much discussed "supervenience argument", which ultimately amounts to the dilemma that mental states either are causally inefficacious or they hold the threat of overdetermining neural states. Forced by this disjunction, Kim votes in favor of overdetermination and, ultimately, reduction.

We propose a perspective on mental causation that dissolves the assumption of a tension between horizontal and vertical determination. For mental states to be causally efficacious, they must be dynamically stable. This important requirement can be implemented by combining a key idea of supervenience, multiple realization, with the recently introduced vertical interlevel relation of contextual emergence. Both together deflate Kim's dilemma and reflate the causal efficacy of mental states.

Keywords: contextual emergence, mental causation, multiple realization, overdetermination, supervenience.

Accepted for publication in:
Journal of Theoretical and Philosophical Psychology

1 Introduction

This article addresses the philosophical puzzle of how the mind can be causally relevant in a physical world: the "problem of mental causation".[1] The question of how mental phenomena can be causes is of high significance for an adequate comprehension of scientific disciplines such as psychology and cognitive neuroscience. Moreover, mental causation is crucial for our everyday understanding of what it means to be an agent in a natural and social environment. The literature on this topic is overwhelming, and we can only address a section of it that, to our knowledge, has been most influential and, for our purposes, is most interesting.

One of the reasons why the causal efficacy of the mental has appeared questionable is that a horizontal determination of a mental event m by prior mental events seems to be inconsistent with a vertical determination of m by neural events. If a vertical neural determiner of m is brought about, then according to supervenience m is secured whether or not its horizontal mental determiner occurred as well. This has been used to argue that reasoning about mental causation leads us into a dilemma: either mental events play no horizontally determining causal role at all (horn 1), or they are causes of the neural bases of their respective horizontal mental effects (horn 2) (cf. Kim 2003).

In this paper, we shall show that the alleged conflict between horizontal and vertical determination of mental events is ill-conceived. Exploiting recent progress in theoretical neuroscience, we show that there are independent reasons against both horns of the mentioned dilemma. The key point, to be explained in detail below, will be that mental states (related to but not identical with the philosophical notion of mental events) can be properly constructed from the dynamics of an underlying neural system. This gives rise to a mental dynamics independent of those neurodynamical details that are irrelevant for a proper construction of mental states.

The horizontal determination relation connecting mental states is based on their dynamics and satisfies certain counterfactual criteria that are strongly indicative for causation. The same criteria are not satisfied, however, by individual mental states with respect to those individual neural states that underlie their horizontal mental effects. As a consequence, we argue that (i) mental states can indeed be causally and diachronically related to other mental states (falsity of horn 1), and (ii) they are neither causally related to their synchronic neural determiners nor to the neural determiners of their horizontal effects (falsity of horn 2). This makes a strong case against a conflict between a horizontal and a vertical determination of mental events and resolves the problem of mental causation in a deflationary manner. Vertical and horizontal determination do not compete, but complement one another.

The article is organized as follows. Sec. 2 contains a detailed presentation of the problem of mental causation within the framework of the so-called "supervenience

[1] For an extensive review of a range of solutions to the problem, see Robb and Heil 2009. For a detailed exposition of the different versions of the problem see Harbecke 2008, ch. 1.

argument". The first stage of this argument establishes the mentioned dilemma of the two horns and serves as the basis for our argumentation.

In Sec. 3 we address relations between the mental and the neural in terms of a recently introduced interlevel relation denoted as "contextual emergence" (Bishop and Atmanspacher 2006; Atmanspacher and beim Graben 2007; Atmanspacher 2009).[2] In this framework we combine particular details pertaining to supervenience with a particular kind of vertical (synchronic) emergence. The necessary mathematical tools for this supervenience-based emergence are subtle and lead to a theoretically sound and empirically applicable relation between mental and neural states. A key point of this approach is its efficient exploitation of the temporal evolution (dynamics) of neural states for the construction of properly defined mental states.

Sec. 4 summarizes and compares the essential notions of Secs. 2 and 3. Finally, in Sec. 5, we utilize the ideas of contextual emergence to propose an alternative view of mental causation. We argue that the two horns of the dilemma in the supervenience argument are unfounded and present a constructive alternative that retains the option that mental states can be causally efficacious. Sec. 6 summarizes our results. Our conclusions match with and refine the notion of "proportionate causation" introduced by Yablo (1992).

As a methodological point, our approach links theoretical and empirical results of cognitive neuroscience, in particular cognitive neuro*dynamics*, with philosophical arguments about mental causation, a key ingredient of our role as agents in the world. Both the philosophy of science and the philosophy of mind need such connections if they are interested in the details of whether and how their arguments and conclusions relate to scientific results. Our work also exemplifies that a particular focus on neuro*dynamics* has the potential to help understanding traditional problems in philosophy in a novel framework.

2 Mental Causation and the Supervenience Argument

2.1 The Problem of Mental Causation

The problem of mental causation[3] is a comparably recent problem in the history of philosophy. It started to become gradually recognized as a genuine problem when "physico-mechanical" thinking began to have an influence in philosophy, notably through the works of René Descartes with his dualist ontology of thinking substance (*res cogitans*) and extended substance (*res extensa*). The question, posed by

[2] The technical notion of contextual emergence differs from classical ideas of emergence as promoted by the "British emergentists" and their successors (cf. McLaughlin 1992).

[3] For a more comprehensive description of the history of the problem of mental causation cf. Harbecke 2008, ch. 1.

Elizabeth of Bohemia in 1643, of how "man's soul (thinking substance) can determine animal spirits so as to cause voluntary actions" (Adam and Tannery 1904, III: 661) led Descartes in his *Passions of the Soul* of 1649 (Adam and Tannery 1904, XI) to his infamous hypothesis that the pineal gland is responsible for mediating between the two substances.

In the centuries following Descartes, the problem of mental causation developed into a widely discussed issue in philosophy. Virtually all prominent philosophers of subsequent centuries proposed solutions to it.[4] Nevertheless, it was not before the second half of the twentieth century that the problem of mental causation moved into mainstream metaphysics, where it has remained until today (cf. Kim 1997). After C.D. Broad's classic *The Mind and Its Place in Nature* of 1925, Ludwig Wittgenstein's reflections on mentality and mental discourse, and Gilbert Ryle's *The Concept of Mind* (1949), the first canonical formulations of the problem are due to Place (1956), Feigl (1958), and Smart (1959) in the late 1950s.

The distinctive feature of the modern version of the problem consists in its explicit concern with the metaphysical status of mental phenomena rather than with the logic of psychological explanation. The positions on the status of the mind in a physical world discussed by Place, Feigl, and Smart already displayed interesting commonalities with today's debate. However, the specific form in which the problem of mental causation is discussed today was proposed in the late 1960s by authors such as Malcolm (1968) and Goldman (1969) who construed the puzzle as a particular set of seemingly incompatible assumptions.

The debate in the 1970s was dominated by functionalist approaches to the mind (cf. Putnam 1967, 1975; Fodor 1974), which was initially believed to evade the problem of mental causation by considering mental phenomena as functional phenomena. After a while, however, it became clear that functionalism does not provide a satisfactory solution, mainly through the work of Kim (1979, 1985), who has long aimed for a maximally transparent statement of the problem. Since then, the literature has mushroomed.[5]

Current discussion of the problem of mental causation usually frames the problem in terms of mental and physical *events* involving properties, so that the problem is framed within a property-dualist rather than a substance-dualist perspective. The problem is usually formulated by a set of four assumptions (A1)–(A4) each one of which appears plausible in isolation. However, the conjunction of these assumptions is provably inconsistent.

---

[4]For some selected examples, cf. Leibniz (1890, 6), Spinoza (1677, sec. 2.1.2), Kant (1910, 1:19-21), Schopenhauer (1980, 3:171-173), and Huxley (1893).

[5]For a small collection of authors representing the spectrum of positions developed in this late stage of the debate on mental causation, cf. Bennett (2003), Bishop (2006) Block (1997, 2003), Bontly (2002), Campbell (1970), Crisp and Warfield (2001), Dardis (2008), Elder (2001), Fodor (1997), Harbecke (2008), Kim (1998, 2003, 2005), Loewer (2007), MacDonald and MacDonald (1986), Marras (2007), Noordhof (1999), Papineau (2002), Schiffer (1987), Sosa (1984), Yablo (1992).

(A1) Some mental events cause physical events.
(A2) Every physical event has a complete sufficient physical cause.
(A3) Mental events are not identical with physical events.
(A4) Physical events are not pervasively causally overdetermined.

The notion of an "event" figuring in these assumptions is usually interpreted in a "fine-grained" or "intensional" way (Kim 1973). Events, according to this interpretation, are instantiations of a property by an individual at a time. Two events are identical if (i) the objects, (ii) the properties, and (iii) the times figuring in them are identical.[6] Modelling events in this way is motivated by the intuition that events have their effects not *simpliciter* but always *in virtue of* some or all of their properties (cf. Honderich 1982).[7] In this sense, a mental event is understood as the instantiation of a mental property by some individual at a time, and a physical event is understood as the instantiation of a physical property by some individual at a time.

Mental events appear to play an indispensable role when we interpret actions of agents. Assuming that mental events never cause anything would amount to the contention that virtually all our interpretive practices are pointless. The inacceptability of this consequence suggests (A1). Any physical event, on the other hand, typically seems to have a complete sufficient physical cause, suggesting (A2).[8]

The conjunction of assumptions (A1) and (A2) implies that mental events must (i) either be redundant causes of physical events or (ii) they must themselves be physical events. The latter option has turned out difficult to defend within the fine-grained model of events. According to this model, for two events to be identical it is necessary that the properties of the events be identical. However, since few paradigm mental properties are coextensional with a (natural and non-disjunctive) physical property (Putnam 1967) and since coextensionality is necessary for identity, it has proven problematic to justify an identity claim with respect to mental and physical properties in almost all interesting cases. Hence, the events in which mental

[6] We deliberately leave implicit what precisely individuals, properties, and times are. However, a minimal interpretation of individuals could be three-dimensional space regions, properties could be ascribed to certain sets of individuals, and times could be intervals on the time-axis of a four-dimensional universe. Objects can be thought of as sequences of events. In Sec. 3 we will use a different terminology, adopted from the physics of dynamical systems, and relate it to the philosophical term of an "event" in Sec. 4.3. In particular, whilst Kim construes events independently of context, in our approach the sameness of context is significant for their identity.

[7] The counterpart underpinned by an extensional intuition of causation would be a "coarse-grained" model of events, cf. Davidson 2001, Appendix B, 309.

[8] It should be noted, however, that the doctrine of the causal completeness of the physical is less obvious than its aquiescent acceptance in the philosophical literature suggests (cf. Montero 2003; Primas 2007). Bishop and Atmanspacher (2011) have argued that the notion of causation is inconsistent with the fundamental laws of physics insofar as these laws have no direction of time, hence no past and future, hence no cause and effect. This inconsistency is easily ignored but not easy to dispel.

properties figure turn out to be not identical with any physical event. This supports assumption (A3).

If mental events are not identical with physical events, then only the first among the two options (i) and (ii) above remains: They can only be redundant causes of physical events. This, however, contradicts assumption (A4), so that the full set of assumptions (A1)–(A4) is inconsistent. The reason for holding on to (A4) is that redundant causes *per definitionem* do not make any difference to what goes on in the world, and it seems extravagant to assume the existence of things that make no causal difference. This suggests (A4).

Most attempts to resolve the problem have focused on the question which of the assumptions can be dropped at the lowest overall theoretical costs. Any choice for the rejection of one of the assumptions forms the basis of a family of solutions to the problem (cf. Harbecke 2008, ch. 2). In contrast to this general strategy, some authors have pointed out that the problem is invalid in that it equivocates at least one central term. In particular, it has been contended that many mental events do not cause physical events at all but cause further mental events (Gibbons 2006, 99). If assumption (A1) is reformulated in this sense, the inconsistency of (A1)–(A4) disappears. Similarly, it has been suggested that those mental events that are causes of physical events as stated by assumption (A1) cause *macro*physical events exclusively. However, assumption (A2) is convincing only if it claims a causal completeness for the *micro*physical (Sturgeon 1998, 415/16). If the equivocation is based on a non-trivial move from micro- to macrophysical events the inconsistency of (A1)–(A4) disappears again.

2.2 The Supervenience Argument

The line of argument mentioned at the end of the previous section has drawn criticism most notably by Kim (1998, 38-47; 2003; 2005, 13-22, 39-70). As Kim has argued, even if the inconsistency of (A1)–(A4) disappears due to a reformulation of an assumption, a new inconsistency arises once the almost invariably accepted hypothesis of a supervenience of mental events on physical events is taken into account and added as a new premise to assumptions (A1)–(A4). For example[9] if (A1) is reformulated as (A1)′ below, the new problem contains the following assumptions:[10]

(A1)′ Some mental events cause future mental events.
(A2) Every physical event has a complete sufficient physical cause.

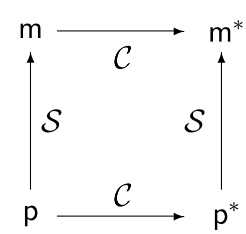
[9] An analogous reformulation can be constructed for the claims that mental events cause macrophysical events and that only the microphysical can be called causally complete; see Harbecke (2010).

[10] This formulation can be mapped onto the formulation used by Kim 2003: Assumption (A1)′ mirrors Kim's premise (1); assumption (A2) mirrors the second conjunct of Kim's premise (5); assumption (A3) mirrors Kim's premise (6); assumption (A4) mirrors Kim's premise (7); assumption (A5) summarizes Kim's premises (2) and (4).

(A3) Mental events are not identical to physical events.
(A4) Physical events are not pervasively causally overdetermined.
(A5) Any mental event has a physical event as its supervenience base.

It is not immediately clear what it means that a mental event m has a physical event p as its supervenience base since virtually all canonical formulations of the concept of supervenience are not about relations between events but rather between classes of properties (cf. McLaughlin and Bennett 2008). However, there is a simple and plausible interpretation of the notion of an event p as a supervenience base of an event m, according to which the individuals and times of p and m are identical, and the property P instantiated by p is sufficient for the property M instantiated by m. This parlance refers to a "strong supervenience" (cf. Kim 1984, Sec. 2.2) of the property class of M on the property class of P.

According to this understanding, the occurrence of p is sufficient for m, but not necessarily *vice versa*. The interpretation would therefore be consistent with both an identity and a non-identity of p and m.[11] Hence, (A5) would not immediately contradict (A3), which is the desired outcome. In what follows we will use the notion of a "vertical" determination of events widely in this sense. The general picture of how mental causation works as described by (A1)′–(A5) is often depicted by the following diagram, or variants of it (cf. LePore and Loewer 1989, 180; Kim 2003, 159; Gibbons 2006, 79; Kallestrup 2006, 463; Sachse 2007, 52; Bennett 2008, 302; m, m*: mental events; p, p*: physical events; $\mathcal{S}$: vertical (synchronic) determination; $\mathcal{C}$: horizontal (diachronic) determination):



Kim believes that this picture is problematic, or at least so, if the arrows are interpreted not only as horizontal/vertical determination but as *direct* horizontal/vertical determination.[12] He identifies an intuitive "tension" (2003, 155) between the claim

[11]Another way of putting this point is to say that the supervenience base p of m may, or may not, be also a reduction base of m. With Kim's criteria of event identity in the background (cf. Kim 1973, p. 223), a supervenience base p is a reduction base of an event m only if the property instantiated by m is identical to the property of p. However, property identity may not hold while p may still be a supervenience base of m.

[12]Kim never makes this distinction explicitly. However, without it the supervenience argument

expressed by the arrow connecting m and m^* and the claim expressed by the arrow connecting p^* and m^*. Kim's intuition seems to be based on a general exclusion principle saying that, rare exceptions neglected, no event can have more than one direct determiner (unless all but one of the direct determiners are identical to the determined event itself, given that identity is a kind of direct determination).[13] Hence, if (A1)′ should really be interpreted as saying that "some mental events are direct determiners of future mental events" as suggested by the arrow connecting m and m^*, then the conjunction of (A3), (A5) and the exclusion principle implies the falsity of (A1)′ (horn 1).

It is important to see, however, that Kim is not content with this conclusion. When he points out that m cannot be considered a determiner of m^*, "… unless … the occurrence of [m] had something to do with the occurrence of [p^*]" (2003, 156), it becomes clear that he takes into consideration a weaker interpretation of (A1)′, namely "some mental events are determiners of future mental events." Under the plausible assumptions that any determiner is either a direct or an indirect determiner, and that any indirect determiner of an event must determine all direct determiners of that event, Kim can infer that m is an indirect determiner of m^* by being a determiner of p^* (horn 2; cf. 2005, 156: assumption (3)). The latter assumption corresponds strongly to premise (A1), which lets Kim conclude that "…'same-level' causation entails 'downward' causation." (2003, 156)

To summarize, Kim first shows that, depending on whether (A1)′ makes a claim about direct horizontal, or just horizontal, determination, the mentioned exclusion principle underpinning the intuition of a "tension" between (A1)′ and (A5) implies that either mental events do not horizontally determine anything (horn 1) or they horizontally determine physical events (horn 2).[14]

Our aim in the following is to provide evidence against both horn 1 and horn 2. We will then show by two *reductio* arguments that the exclusion principle underpinning Kim's intuition of a "tension" between (A1)′ and (A5) comes out inaccurate

becomes unintelligible as Kim eventually does claim that m determines m^* horizontally, namely indirectly through p^*.

[13] One of the puzzling a spects of Kim's argument is that he has explicitly declared it preferrable "…not to appeal to any general principle here; I now prefer to rely on the reader's seeing the tension.… I don't believe invoking any 'principle' will help persuade anyone who is not with me here." (2005, 42n.8) However, if the intuition of a "tension" is to be relevant for a deductive argument (and Kim explicitly speaks of an "entailment"; cf. 2003, 156), then the intuition *must* be describable as a principle-like general assumption. For a detailed defense of this consequence, cf. Harbecke 2010.

[14] It should be pointed out that the dilemma described here results from what Kim characterizes as the "first stage" of the supervenience argument (cf. 2003, 155/156). The ultimate aim of Kim's argument is to show that "either reduction or causal impotence" (2003, 165) must be inferred for mental events. The second disjunct of this conclusion ("causal impotence") is horn 1. The first disjunct ("reduction") is inferred from horn 2 in what Kim describes as the second stage of the argument (cf. 2003, 157-166). In this paper, we are concerned with the first stage only. This is because, if the first stage has been deflated, the second becomes irrelevant.

no matter which of the two mentioned interpretations is given to (A1)′. Hence, we acknowledge that Kim's argument itself is valid, but deny that it is sound because it relies on the inaccurate "tension" premise. Horizontal and vertical determination do not compete, but complement one another. This insight paves the way to a deflationary solution of the problem of mental causation.

3 Relations between Mental and Neural States in Contextual Emergence

3.1 Conceptual Scheme

The basic idea of contextual emergence is that, starting at a particular level L of description of a system, a two-step procedure can be carried out that leads in a systematic and formal way (1) from an individual description L_i of the system to a statistical description L_s of the system and (2) from L_s to an individual "higher-level" description H_i. This scheme can in principle be iterated across any connected set of descriptions, so that it is applicable to any case that can be formulated precisely enough to be a sensible subject of a scientific investigation.

The essential goal of step (1) is the identification of equivalence classes of individual states that are indistinguishable with respect to a particular ensemble property. Insofar as this step implements the multiple realization of statistical states in L_s by individual states in L_i, it is a key feature of a supervenience relation with respect to states.[15] The equivalence classes at L can be regarded as cells of a partition. Each cell can be regarded as the support of a (probability) distribution representing a statistical state.

The issue of composition, which is emphasized in alternative types of emergence, is to be treated in the framework of this step (1). In contextual emergence, however, the point is not the composition of large objects from small ones. Rather than size, the point here is that statistical states are formulated as probability distributions over individual states. This way they can be considered as both compositions and representations of (limited) knowledge about individual states.

The essential goal of step (2) is the assignment of individual states at level H to coextensional statistical states at level L. This cannot be done without additional information about the desired level-H description. In other words, it requires the choice of a context setting the framework for the set of observables (properties) at level H that is to be constructed from level L. The chosen context provides conditions to be implemented as a stability criterion at level L.

This stability criterion guarantees that the statistical states of L_s are based on a robust partition so that the individual states and emergent observables in H_i are

[15] Note that the notion of a state is not identical with the philosophical notion of an event. Their relation will be clarified in Sec. 4.

not ill-defined. For instance, if a partition is not stable under the dynamics of the system at L, the assignment of states in H_i will change over time and is not well-defined in this sense. The implementation of a contingent context of H_i as a stability criterion in L_i yields a proper partition for L_s. In this way, the lower-level state space is endowed with a new, contextual topology, i.e. a coarse graining tailored to the context considered (see Atmanspacher 2007 for more details).

As the lower-level partition can be refined (under the dynamics), it gives rise to a multitude of statistical states of L_s, depending on the degree of refinement. In principle, all cells of the partition, i.e. all corresponding L_s-states, are candidates for an identification as H_i-states. However, the higher-level context decides which refinement within step (1) is relevant for an empirically given or theoretically considered situation.[16] For instance, a highly refined partition is inappropriate for an analysis in which only two higher-level states are involved (requiring a bi-partition).

From a slightly different perspective, the context selected at level H decides which details in L_i are relevant and which are irrelevant for H_i. Differences among all those individual states at L_i that fall into the same equivalence class at L_s are irrelevant for the chosen context. In this sense, the contextually determined partition at L_s is based on both stability and relevance conditions. This interplay of context and stability across levels of description is the core of contextual emergence.

The overall picture yielded by steps (1) and (2) together looks as if sufficient (but not necessary) conditions by supervenience in (1) and necessary (but not sufficient) conditions by emergence in (2) add up to full-blown reduction, where the lower-level description is both necessary and sufficient for the higher-level description in the first place. But this is not the case. The reason is that a higher-level context must be implemented at the lower level in order to get the partition that is in turn required to define states at the higher-level description. Only *after* the higher-level context is implemented as a lower-level stability criterion does the lower-level description indeed carry both necessary and sufficient conditions. In other words, the combination of supervenience and emergence must be *self-consistent* with respect to the selected context. This self-concistency guarantees that two systems with identical lower-level features cannot have different higher-level features under an assumed higher-level context.

As a concrete example for contextual emergence in physics, consider the transition from classical point mechanics over statistical mechanics to thermodynamics (Bishop and Atmanspacher 2006). Step (1) in the discussion above is here the step from point mechanics to statistical mechanics, essentially based on the formation of an ensemble distribution. Particular properties of a many-particle system are defined in terms of a statistical ensemble description (e.g., as moments of a many-particle distribution function) which refers to the statistical state of an ensemble

[16]Similarly, Fodor (1974, p. 113) stated that higher-level properties are not salient from the perspective of lower-level properties: "[E]ven if brains were out where they could be looked *at*,... we wouldn't know what to look *for*" when we develop psychological theory. Contextual emergence gives a detailed picture of why this is so.

(L_s) rather than the individual states of single particles (L_i).

Step (2) is the step from statistical mechanics to thermodynamics. Concerning observables, this is the step from the expectation value of a momentum distribution of a particle ensemble (L_s) to the temperature of the system as a whole (H_i). In many philosophical discussions this step is mischaracterized by the false claim that the thermodynamic temperature of a gas is identical with the mean kinetic energy of the molecules which constitute the gas. In fact, a proper technical discussion of the subtle details showing that this is not the case was not available before Haag (1974) and Takesaki (1970).

This brief survey shows that quite some sophisticated argumentation is required to establish Nagel's innocently looking "bridge law" relating temperature to mean kinetic energy (Nagel 1961, ch. 11). In the debate about reductive accounts, the *derivation* of such relations in this and other examples typically remains disregarded, while contextual emergence suggests a formal option to derive such relations rigorously. This may indicate why philosophers of science have often prematurely embraced an identity claim and, as a consequence, simple reduction. We will come back to this issue in Sec. 5.1 below.

There are other examples in physics and chemistry which can be discussed in terms of contextual emergence: emergence of geometric optics from electrodynamics (Primas 1998), emergence of electrical engineering concepts from electrodynamics (Primas 1998), emergence of chirality as a classical observable from quantum mechanics (Bishop 2005; Bishop and Atmanspacher 2006), emergence of hydrodynamic properties from many-particle theory (Bishop 2008).

3.2 Mental States from Neurodynamics

If descriptions at L and H are well developed (as in the examples just mentioned), a formally precise interlevel relation can be straightforwardly set up. The situation becomes more difficult in situations where no such established descriptions are available. This is the case in certain areas of cognitive neuroscience or consciousness studies focusing on relations between neural and mental states (e.g., the identification of neural correlates of conscious states).[17]

For the application of contextual emergence, the first desideratum is the specification of proper levels L and H. With respect to L, one needs to specify whether states of neurons, of neural assemblies or of the brain as a whole are to be considered; and with respect to H a class of mental states reflecting the situation under study needs to be defined. In a purely theoretical approach this can be tedious, but in empirical investigations the experimental setup can often be used for this purpose. For instance, experimental protocols include a task for subjects that defines possible mental states, and they include procedures to record brain states.

[17]In this and the following sections we refer to the "neural" as a specification of the "physical". The concrete scenarios and applications of contextual emergence in cognitive neuroscience always deal with correlations between neural and mental states.

A theoretical framework for the contextual emergence of mental states from neural states is due to Atmanspacher and beim Graben (2007). A concrete demonstration of how it works for experimental data has been given by Allefeld et al. (2009). Both are based on the so-called state space approach to mental and neural systems, see Fell (2004) for a brief introduction.

The first step is to find a proper assignment of L_i and L_s at the neural level. Possible candidates for L_i are the states and properties of individual neurons. Then the first task is to construct L_s in such a way that statistical states are based on dynamically stable equivalence classes of those individual states whose differences are irrelevant with respect to a given mental state at level H. This reflects that a neural correlate of a conscious mental state can be multiply realized by "minimally sufficient neural subsystems correlated with states of consciousness" (Chalmers 2000).

For neural systems, where complicated dynamics far from thermal equilibrium are involved, a powerful method to implement a given H-context as a L-stability condition uses the neurodynamics itself to find proper statistical states. The essential point is to identify a partition of the neural state space whose cells are robust under the dynamics. This guarantees that individual mental states H_i, defined on the basis of statistical neural states L_s, remain well-defined as the system develops in time. This way, differences between individual neural states L_i belonging to the same statistical state L_s are deliberately disregarded.

For multiple fixed points, their basins of attraction yield a proper partition, while chaotic attractors need to be coarse-grained by so-called generating partitions. Both can be determined from experimental data and provide a rigorous theoretical constraint for the proper definition of stable mental states. The formal tools for the mathematical procedure derive from the field of symbolic dynamics (see Appendix for details), and are discussed and applied in Atmanspacher and beim Graben 2007 and Allefeld et al. 2009.

Let us emphasize again that the neural *dynamics* is absolutely essential for the proper identification of equivalence classes, because well-defined mental states multiply realized by neural states must be dynamically stable. This dynamical stability is guaranteed if the dynamics maps each point on a boundary between equivalence classes (or partition cells) onto another point on a boundary. The concept of a generating partition, introduced by Kolmogorov (1958) and Sinai (1959), provides us exactly with such a construction (for more details and definitions see the Appendix and Atmanspacher and beim Graben 2007).

If the dynamics of neural states p is Φ and the dynamics of mental states m is Γ, then Φ and Γ are related by

$$\pi \circ \Phi = \Gamma \circ \pi \, ,$$

where π is a mapping from the neural state space to the mental state space. The vertical and horizontal relations in this picture can be represented diagrammatically as:

$$
\begin{array}{ccc}
m & \xrightarrow{\Gamma} & \Gamma(m) \\
\uparrow \pi & & \uparrow \pi \\
p & \xrightarrow{\Phi} & \Phi(p)
\end{array}
$$

Although this diagram looks deceptively similar to that of Sec. 2.2, according to the supervenience argument, there are a number of subtle but important differences to be highlighted in Sec.4.3.

As explained in the appendix, the intertwiner π preserves the topology of the neural state space if the partition yielding the equivalence classes of neural states is generating. For generating partitions of the neural state space, the neural dynamics Φ and the mental dynamics Γ are said to be "topologically equivalent", or: Γ is a "faithful representation" of Φ (see the Appendix for brief characterizations of these technical terms).

This implies a one-to-one correspondence between statistical neural states and individual mental states, precluding any inconsistencies, tensions, or competitions between the two. Horizontal causal relations are contained in Γ and Φ, while vertical interlevel relations are contained in π. In this picture, multiple realization features as a non-causal intralevel relation transforming individual states into statistical states at the neural level.

3.3 A Pertinent Example

A pertinent example for the application of contextual emergence to experimental data is the relation between mental states and electroencephalographic (EEG) states. In a recent study, Allefeld et al. (2009) tested the method using data from the EEG of subjects with sporadic epileptic seizures. This means that the neural level is characterized by brain states recorded via EEG, while the context of normal and epileptic mental states essentially requires a bipartition of that neural state space into two cells.

The analytic procedure starts with a (for instance) 20-channel EEG recording, giving rise to a state space of dimension 20, which can be reduced to a lower number by restricting to principal components. On this state space, a homogeneous grid of cells is imposed as a fine-grained auxiliary partition to set up a transition matrix reflecting the EEG dynamics.

The eigenvalues of this matrix yield time scales for the dynamics which can be ordered by size. Gaps between successive time scales indicate groups of eigenvalues defining partitions of increasing refinement (Allefeld and Bialonski 2007). In the case considered, there is a significant gap after the second largest eigenvalue, indicating that two equivalence classes of neural states are appropriate for the desired partition. In other words, the context of two mental states investigated (normal versus seizure) must be implemented by two dynamically stable partition cells in the neural state space.

This can be achieved by looking at the eigenvectors corresponding to the first two eigenvalues. They span an eigenvector space in which the empirically obtained data points form a 2-simplex. Each data point can now be classified according to that vertex of the simplex to which its distance is closest. This classification allows us to determine which data point belongs to which partition cell. All points closer to vertex 1 will be assigned to mental state 1 (normal), and all points closer to vertex 2 will be assigned to mental state 2 (seizure).

Finally, the result of the partitioning can be inspected in the originally recorded time series to check whether mental states are correctly related to the proper episodes in the EEG dynamics. The study by Allefeld et al. (2009) shows perfect agreement between the distinction of normal and epileptic states and the bipartition resulting from the spectral analysis of the neural transition matrix, i.e. the neural dynamics.[18]

The generating partition found entails that the intertwiner π in the diagram in Sec. 3.2 indeed gives rise to (i) the coextensionality of statistical neural states and individual mental states and (ii) the topological equivalence of neural and mental dynamics. As we show below, these results are directly relevant for a proper assessment of the dialectics of the supervenience argument presented in Sec. 2.2.

4 Terminological Clarifications

Kim's formulation of the problem of mental causation and the way in which contextual emergence addresses this problem differ conceptually and terminologically. In order to identify the conceptual differences, let us first recapitulate the central notions of both approaches in a compact way and relate them to one another.

4.1 Key Terms of the Supervenience Argument

As mentioned in Sec. 2.1, an event in Kim's terminology is characterized as a set $\{(x,t),A\}$, where x,t is the definite spatiotemporal location of the event and A is a property (Kim 1973, 222).[19] An event is called a mental event m if the property belongs to a mental property class; it is called a physical (neural) event p if the property belongs to a physical (neural) property class.

The horizontal relation $\mathcal{C}$ relates a previous mental event m (at t_1) to a future mental event m^* (at t_2), and a previous neural event p (at t_1) to a future neural

[18] Another recent application of contextual emergence refers to the relation between lower-level neural (micro-) states (e.g.,individual neurons) and higher-level neural (macro-) states (e.g., functionally coupled neuronal assemblies). Amari and colleagues (Amari 1974, Amari *et al.* 1977) proposed to identify neural macrostates based on two criteria: (i) the structural stability of microstates as a necessary lower-level condition, and (ii) the decorrelation of microstates as a sufficient higher-level condition. The required macrostate criteria are purely stochastic, however, and do not exploit the dynamics of the system in the direct way which a Markov partition allows. A detailed view on Amari's approach in the light of contextual emergence is due to beim Graben *et al.* (2009).

[19] Since Kim includes relations as well, the original notation is in fact "$\{(x_1,\ldots,x_n,t),A^n\}$".

event p^* (at t_2). In both cases, $\mathcal{C}$ refers to "singular causation", due to which two events e_1 and e_2 are related "in virtue of e_1's property E_1".

The "in virtue of" locution indicates that the singular causal relation is backed up by a general causal regularity holding between the property E_1 of e_1 and the property E_2 of e_2. This general causal regularity is only indirectly referred to by the supervenience argument. According to its initial assumptions it holds between the property P of p and the property P^* of p^*, and between the property M of m and the property M^* of m^*.[20]

The vertical relation $\mathcal{S}$ characterizes neural events p and p^* at t_1 and t_2 as the supervenience bases of mental events m and m^* at t_1 and t_2. As mentioned in Sec. 2.2, a plausible way to conceive an event e_1 as a supervenience base of e_2 is to presuppose an identity of the individuals and times of e_1 and e_2, and to presuppose the property E_1 instantiated by e_1 as sufficient for the property E_2 instantiated by e_2 in the sense of a supervenience of E_2's property class on E_1's property class. In other words, the relation of $\mathcal{S}$ is best thought of as backed up by a general sufficiency relation between the property E_1 of e_1 and the property E_2 of e_2 in the sense of property supervenience.

Note that this general sufficiency relation is not symmetric but directed. This is essential for the multiple realization claim that forms the basis for assumption (A3) in Sec. 2.1 and 2.2. According to this claim, neural properties P and P^* of the neural events p and p^* are sufficient, but not necessary, for mental properties M and M^* instantiated by the mental events m and m^*, respectively.

4.2 Key Terms of Contextual Emergence

The scheme of contextual emergence is formally based in the theory of dynamical systems, where the most important concepts are those of states and properties (observables), which undergo a temporal evolution called a dynamics. Individual states x are represented as pointwise elements of a state space X whose coordinates are the properties $\{\mathrm{A}\}$ associated with x. For simplicity, statistical states can be represented by the subsets of X that support their probability measure.

The technique for establishing such subsets, or equivalence classes, of individual neural states in order to define proper statistical neural states p_s is the formal core of contextual emergence. As discussed in Sec. 3 this is done by constructing a partition (a coarse graining) of X in order to identify statistical states p_s that are stable under the neural dynamics Φ. A state p_s is stable under Φ if its boundaries are robust, i.e. Φ maps each point on a boundary onto a point on a boundary. Only if boundary points are mapped onto each other, the bounded subsets of the neural state space, representing statistical neural states p_s, remain intact under Φ. In short, Φ entails a particular partition, the generating partition, that leads to properly

[20]For more details on the distinction between singular and general causation, cf. Baumgartner 2008, 329.

defined, dynamically stable states p_s – all other partitions lead to improperly defined, dynamically unstable states.

The intertwiner π specifies the relation between statistical neural states p_s and individual mental states m_i evolving according to a mental dynamics Γ. Since π is invertible for generating partitions, the diagram in Sec. 3.2 is commutative and we can represent the mental dynamics Γ as the concatenation $\pi \circ \Phi \circ \pi^{-1}$. This expresses that Γ is topologically equivalent with Φ and, thus, a faithful representation of Φ. For generating partitions, the two dynamics $\Gamma(m)$ and $\Phi(p)$ are consistent up to those details of neural states that are irrelevant for the proper definition of mental states.

It is important to keep in mind that neural states are associated with neural properties. The move from an individual to a statistical neural description does not change these properties – it only changes the state concept by adding probability over the neural state space. By contrast, the step from the neural to the mental changes both states and properties. It transforms a statistical neural state p_s into an individual mental state m_i. Since the coordinates of the two state spaces are basically the properties, novel properties unavailable at the neural level are required for the mental level.

4.3 Comparison

The concept of events is lacking in a state space picture (and not a formal part of virtually all areas of physics except relativity theory) and the concept of a state is absent in Kim's discussion (and in most of the philosophy of science and mind altogether). Therefore, the notion of an event as used in Kim's framework must be carefully distinguished from that of a state as used in the framework of contextual emergence. A plausible way to connect the two with each other is the following.

A state x in a state space X serves as a general representation of the states of many specific systems of the same kind (e.g., pendulums, billiards, or else) with their specific spatiotemporal locations in the world. In this sense, a state x in the sense of contextual emergence delineates a particular set of actual (and possible) instances in which a relevant system satisfies a complex predicate involving the properties $\{\mathrm{A}\}$ associated with x and their dynamical evolution.

An event in Kim's sense has a definite location and a definite time, and events are typically subsumed into classes according to their properties. If a state in the state space picture characterizes a particular set of actual (and possible) instantiations of a system, then it is plausible to consider these instantiations as events. As properties according to Kim characterize a particular set of actual (and possible) events, they play a role for events in the supervenience argument that looks similar to the role of states for what events could be in the state space picture.

However, due to the intricate relations between these concepts, there are various risks of confusion. For instance, the two dynamics Φ and Γ involve the notion of time as a parameter. It serves no more than the purpose of smaller-greater distinctions ($t_1 < t_2$ or $t_1 > t_2$), such as in "if a system was in state x_1 at t_1, it will be in state x_2 at t_2". In particular, parameter time does not permit indexations such as "system S located in lab 2.344 at Tel Aviv University was in state x_1 at 13:37 before it switched to state x_2 at 14:05". Thus, the dynamics Φ and Γ provide horizontal regularity statements while $\mathcal{C}$ between events is used in horizontal singular statements. Similarly, the intertwiner π represents a general vertical relation between states rather than the singular vertical supervenience relation $\mathcal{S}$ between events.

While the neural properties P and P^* are presupposed as sufficient, but not necessary, for properties M and M^* respectively, the analysis of Sec. 3.2 describes the states related by π as both necessary and sufficient for one another. This important difference results from the fact that the role of supervenience in the supervenience argument refers to neural and mental events (and their properties), while contextual emergence utilizes multiple realization simply for the construction of statistical neural states as equivalence classes of individual neural states.

Thus, as a first point, P and P^* cannot be identified with states p and $\Phi(p)$. Rather, p and $\Phi(p)$ correspond to particular "broad-grained" characterizations of events (cf. Bechtel and Mundale 1999) at best, to which the supervenience argument makes no explicit reference. As a second point π provides a symmetric picture crucial for the coextensionality of p_s with m_i and the topological equivalence of Φ and Γ as oppposed to the directed picture in event supervenience. The relation $\mathcal{S}$ would correspond to specific instantiations of π in the sense of singular event supervenience.

supervenience argument		contextual emergence
mental events m	instantiate	mental states m_i
neural events p	instantiate	neural states p_i
coarse neural events	instantiate	neural states p_s
singular mental causation $\mathcal{C}$	instantiates	mental dynamics Γ
singular neural causation $\mathcal{C}$	instantiates	neural dynamics Φ
event supervenience sufficiency $\mathcal{S}$	instantiates	intertwiner π

Table 1: Comparison between events and their relations to one another in the supervenience argument and states and their relations to one another in contextual emergence. More details and additional correspondences are explained in the text.

The correspondences listed in Table 1 imply that the two diagrams in Sec. 2.2 and 3.2, although they look strikingly similar, connect different notions. The former illustrates singular relationships between events as instantiations of states, whereas the latter illustrates general relationships between states. It also makes clear that what we refer to as horizontal and vertical determination actually splits up into two pairs of relations: in singular horizontal and vertical determination and in general horizontal and vertical determination.

All these relations can either explicitly or implicitly be located in the supervenience argument and in contextual emergence. This makes it possible to relate the two frameworks to one another in a sufficiently precise way. As a consequence, we will subsequently (in Sec. 5) use particular aspects of the framework of contextual emergence to attain a better understanding of both the supervenience argument and the problem of mental causation.

5 Dissolving the Problem of Mental Causation

5.1 The *Non-Sequitur* of the Supervenience Argument

The invertibility of the intertwiner π in contextual emergence implies that statistical neural states map one-to-one onto individual mental states. One may want to interpret this result as lending support to a hypothesis promoted by Bechtel and Mundale (1999) according to which "broad-grained" neural properties often map one-to-one onto particular mental ("psychological") properties. Such a coextension of mental and broad-grained neural properties contradicts the multiple realization of mental properties. This challenges premise (A3) presented in Sec. 2.1, which is mainly based on the multiple realization of mental properties. If premise (A3) is false, the supervenience argument becomes irrelevant and the problem of mental causation receives a straightforward solution.

But such a reasoning is problematic and it does not revoke the supervenience argument. Within the framework of contextual emergence none of the mental states m_i (and, hence, none of the statistical neural states p_s) is reducible to, or identical with, any individual neural state p_i. If a system is in an individual mental state m_i (and, hence, in a statistical neural state p_s), it is also in a particular individual neural state p_i which is sufficient for both p_s and m_i. However, m_i is not coextensional with p_i, blocking a reduction of m_i (and p_s) onto p_i. This shows that a system's being in mental state m_i (and p_s) is not identical to its being in any individual neural state p_i that realizes p_s (in the sense of multiple realization).

In the language of the supervenience argument, this means that a mental event m instantiating a mental property M is typically not identical to any neural event p instantiating a "narrow-grained" neural property $\mathsf{P_i}$ that is sufficient, but not necessary, for M. Read this way, premise (A3) is strongly supported by contextual emergence and cannot be dismissed. Hence, contextual emergence differs from Bechtel and Mundale's proposal which, as a consequence, cannot be utilized for our solution of the problem of mental causation.

Nevertheless, the formally sound and empirically applicable construction of mental states à la contextual emergence can be shown to take the wind ouf of the sails of the supervenience argument. The point is that contextual emergence provides reasons for both holding onto all premises (A1)′-(A5) and for denying that there is any substantial "tension" between premises (A1)′ and (A5) which could be used to infer (A1). Or, more precisely, contextual emergence gives strong reasons to deny the assumption capturing Kim's intuition of a "tension" between premises (A1)′ and (A5).

To see this, note that contextual emergence not only allows us to construct statistical neural states that map one-to-one onto mental states, but it also provides a faithful representation of the neural dynamics Φ in terms of a mental dynamics Γ. The mental dynamics is therefore as *real* as the neural dynamics. Restricting ourselves (for reasons of simplicity) to deterministic dynamics, an individual mental state m_i is sufficient for any successor state $\Gamma(m_i)$, under the mental dynamics Γ, in the sense of a causal horizontal determination. In other words, if a system implements the mental dynamics Γ and if it instantiates a mental state m_i at a time t_1, the system predictably instantiates mental state $\Gamma(m_i)$ at any subsequent time $t_2 > t_1$. To say that the dynamics Γ faithfully represents the dynamics Φ means that both dynamics are topologically equivalent (see Appendix).

If the regularity $m_i \rightarrow \Gamma(m_i)$ is robust (in the relevant circumstances), an instantiation of an individual state m_i, i.e. a mental event m, is sufficient for an instantiation of $\Gamma(m_i)$, i.e. a mental event m^*, *in virtue of* being an instantiation of m_i. This could be modally rephrased by saying "had the instantiation of m_i occurred, the instantiation of $\Gamma(m_i)$ would have occurred as well" (counterfactual sufficiency). Furthermore, any such instantiation of m_i, i.e. a mental event m, and a horizontally determined instantiation of $\Gamma(m_i)$, i.e. a mental event m^*, supports counterfactuals of the form "had the instantiation of m_i not occurred, the instantiation of $\Gamma(m_i)$ would not have occurred either" (counterfactual necessity).

One important argument in favor of such counterfactuals is that, if m_i and its coextensional p_s had not been instantiated, none of the individual neural states p_i sufficient for m_i and p_s could have been instantiated. However, if an instantiation of any such individual neural state p_i had not occurred, then those instantiations of $\Phi(p_i)$ which are horizontally determined by the instantiation of p_i due to the dynamics Φ would also not have occurred. And if an instantiation of $\Phi(p_i)$ had not occurred, then arguably the instantiation of a statistical state $\Phi(p_s)$ realized by $\Phi(p_i)$ would not have occurred either.[21] If we then assume, as before, that $\Phi(p_s)$ is coextensional with $\Gamma(m_i)$, we can see why the instantiation of $\Gamma(m_i)$ would also not have occurred. Hence, the robust regularity of $m_i \rightarrow \Gamma(m_i)$ suggests the correctness of the necessity counterfactual. And since satisfaction of these counterfactuals is widely considered as sufficient for causation[22], their truth provides strong evidence that m_i and $\Gamma(m_i)$ are directly causally related.

[21] Here we disregard cases in which the instantiation of $\Phi(p_i)$ happens to have additional horizontal determiners unconnected to the instantiation of p_i, and cases in which the instantiation of $\Phi(p_s)$ happens to have additional realizers unconnected to the instantiation of $\Phi(p_i)$.

[22] Unfortunately, within the debate on mental causation, the theories of causation presupposed are rarely made explicit. However, virtually all causal theories characterize a satisfaction of the mentioned counterfactuals by two non-overlapping events as strongly indicative for causation, even if they differ in their interpretations as to what makes these counterfactuals true. Consequently, it is enough for us to show that the confirmed mental dynamics implies precisely these counterfactuals with respect to the mental states m_i and $\Gamma(m_i)$ as well as the "direct" connection between them. By contrast, m_i and $\Phi(p_i)$ do not satisfy these conditions.

The gist of this argumentation is that the horizontal and vertical determination of mental and neural states are embedded in an overall systematic framework (cf. the mapping diagram in Sec. 3.2). While Kim's intuitive notion of a tension assumes vertical and horizontal determination as basically independent, we argue that this is not the case. Given a proper definition of dynamically robust mental states as faithful representations of the neural dynamics, no conflict of this vertical relation with the horizontal mental or neural dynamics is possible.

It is important to realize that these conclusions are independent of the question whether the coextensionality of mental and statistical neural states amounts to an identity or not. All that is required for the falsity of horn 1 is that a genuine causal relation holds between an instantiation of m_i and an instantiation of $\Gamma(m_i)$, which is not jeopardized if identity holds. But independent of this point, here are four basic reasons for us to favor coextensionality over identity.

(1) The structure of the mental state space and that of the statistical neural state space are radically different: one of them is an abstract space of symbol sequences and the other one is a probability space over that vector space. (2) Accordingly, individual mental states are represented pointwise, while statistical states are distributions. (3) The types of dynamics of the corresponding states are topologically equivalent, but this is weaker than identical (see Appendix). The neural dynamics is continuous and analog, the mental dynamics is discrete and digital. (4) And, finally, mental and neural observables fall into different algebras with (generally) very different mathematical structure.

An observation lending further support to the claim about a genuine causal relation is that contextual emergence suggests the instantiation of a state $\Gamma(m_i)$, i.e. an event m^*, to be caused neither by an instantiation of a state p_i, i.e. an event p, nor by an instantiation of a state $\Phi(p_i)$, i.e. an event p^*, even if the latter two are sufficient for the former. Since p^* and m^* occur simultaneously but causation requires temporal sequence, p^* and m^* cannot be causally related.

But also p cannot plausibly be considered a cause of m^* – no particular state p_i ever directly and dynamically leads to a relevant $\Gamma(m_i)$. There is always a "conceptual" transition (a partitioning, or coarse graining) involved from p_i to p_s, and there is always a "vertical" transition (an interlevel mapping) involved from p_s to m_i. Both these transitions are not temporally sequential, hence it is implausible, if not meaningless, to refer to them as causal relations (in the sense of efficient causation), and so it is for events as instantiations of states.

If, with this background, an event m (i.e. an instantiation of a state m_i) satisfies certain conditions definitional to causation relative to an event m^* whereas no other event occurring in the general model is a candidate cause for m^*, there are strong reasons to believe that Kim's horn 1 rejecting a causal relation between two events m and m^* is false. Mental events are not causally inefficacious in general but can be genuine causes of future mental events. Since the validity of Kim's argument itself is not in question, one of his premises leading to horn 1 must be false.

To see how contextual emergence also casts doubt on horn 2, suppose that an individual neural state p_i is a state vertically sufficient for an individual mental state m_i in the sense of forming the base of a statistical neural state p_s that is π-coextensional with m_i. Then it can be shown that m_i does not horizontally determine $\Phi(p_i)$. The reason is that the sufficiency of p_i for p_s (which is coextensional with m_i) is asymmetric. Since there are always many neural states p_i realizing a neural state p_s, m_i is not vertically sufficient for any particular state p_i that is definitional to p_s. This in turn implies that m_i is also not horizontally sufficient for any particular state $\Phi(p_i)$ horizontally determined by the states in the cell definitional to p_s.[23]

Hence, m_i is not horizontally sufficient for $\Phi(p_i)$. Since, according to the notion of causation introduced in Sec. 1, such a horizontal sufficiency must hold if an instantiation of m_i (i.e., a relevant mental event m) should be considered a genuine cause of an instantiation of $\Phi(p_i)$ (i.e., a relevant neural event p^*), it is highly doubtful that an instantiation of m_i can be a cause of an instantiation of $\Phi(p_i)$, even if the former is connected to the latter in various other ways. Again, since the validity of Kim's argument is not in question, one of the premises used to infer horn 2 must be false.

Since the rejection of (A1)′ is no longer an option, the prime suspect for an inaccurate premise is not one among the original premises of the supervenience argument. Rather, it is the assumption underpinning the intuition of a "tension" between (A1)′ and (A5). The analytical soundness and empirical adequacy of the framework of contextual emergence provide strong evidence that the assumption of a general exclusion principle for horizontal and vertical determiners is misguided.

Once the corresponding assumption of a "tension" is rejected, premises (A1)′-(A5) can reasonably be considered correct and the conflict between them dissolves. The resulting solution to the problem of mental causation is deflationary: There is no problem with the causal efficacy of mental states on future mental states in the first place.[24]

5.2 Brief Excursion: Proportionate Causation

Finally, let us briefly point out that our results vindicate a now widely received theory most notably defended by Yablo (1992) in important respects.[25] Yablo formulates a set of conditions that jointly define what the author calls "proportionate causation" (*op. cit.*, 273-277). Many events may be causally sufficient, or causally necessary, for a given event, but only those satisfying the conditions definitional to causal proportionality are genuine causes of the event. In many cases, mental or

[23] The pathological case in which all states p_i definitional to a state p_s horizontally determine the same individual neural state under Φ can be excluded.

[24] The question of whether mental states can have causal impact on neural states remains to be clarified in further work.

[25] Yablo's framework has inspired a number of other accounts with a similar structure (cf. Thomasson 1998; Crisp and Warfield 2001; Shoemaker 2001; Gibbons 2006; Schröder 2007).

other higher-level events qualify as genuine proportionate causes of certain effects (*op. cit.*, 277-279).

The distinction of causal sufficiency, causal necessity, and causation is mirrored by the framework of contextual emergence. An instantiation of an individual mental state m_i may be causally necessary for an instantiation of a state $\Phi(p_i)$, in the sense that p_i is an individual neural state vertically sufficient for m_i and horizontally sufficient for $\Phi(p_i)$. However, as argued above, the instantiation of m_i does not satisfy certain other essential criteria of causation with respect to the instantiation of $\Phi(p_i)$ and so cannot be considered a cause of the latter. Nevertheless, an instantiation of p_i is clearly the cause of an instantiation of $\Phi(p_i)$, and an instantiation of m_i is clearly the cause of an instantiation of $\Gamma(m_i)$. This kind of horizontally parallel structure has been shown to be implied by Yablo's account as well (Harbecke 2008, 305/306).

Moreover, Yablo's framework contains the idea that mental and physical events are related by vertical "determination" relations resembling the relations that individual neural states bear to mental states via statistical neural states. Individual neural states are intimately related to statistical neural states, even if the two kinds of neural states are not identical. Determination à la Yablo captures a comparable idea of an intimate relation of determinables to their determinates that is not the identity relation (cf. Yablo 1992, 256-260 and n.29). Since the framework of contextual emergence is clearly more explicit and refined theoretically, and since it has been proven to be successfully applicable to concrete empirical situations, our approach to mental causation advances Yablo's. As an additional benefit, we need no metaphysically problematic notions such as the "essence" of an event (cf. Yablo 1992, 261-265).

6 Conclusions

We discussed the issue of mental causation from the novel perspective of a particular interlevel relation called contextual emergence. It allows us to assess relations between mental features and neural features in a conceptually sound, formally rigorous, and empirically accessible fashion.

We reformulated the problem of mental causation as it arises in the supervenience argument of Kim and others in the framework of contextual emergence. The first crucial point is that mental and neural states, represented in corresponding state spaces in the framework of contextual emergence, are instantiated by mental and neural events as used in the supervenience argument.

This has consequences for the horizontal (intralevel) determination of states / events and for their vertical (interlevel) determination. The relations of causation (horizontal) and supervenience (vertical) between mental and neural events in the supervenience argument are backed up by the dynamics of states (horizontal) and the actual emergence relation (vertical) formalized by an intertwiner between coextensional mental and neural states.

Characterizing mental-neural relations by intertwiners between coextensional states may sound unfamiliar in the terminology of the philosophy of mind. Within the usual concepts of relationships discussed in the literature, the picture that contextual emergence leads to is most closely related to the family of dual-aspect approaches. For different variants of such approaches and their delineation from neutral monism see Atmanspacher (2012).

A careful analysis of the commonalities and differences between the supervenience argument and contextual emergence yields that the problem of mental causation as posed in the supervenience argument does not arise if the appropriate relations in contextual emergence are utilized. The key point is the explicit consideration of the temporal dynamics of states, which is the conceptual basis of any horizontal causation in general. In contextual emergence, the neurodynamics is used to construct statistical neural states which are in one-to-one correspondence with properly defined mental states. Their dynamics is then topologically equivalent with the neurodynamics.

We showed in detail how the framework of contextual emergence suggests the falsity of the two horns of the dilemma established by the supervenience argument. In particular, we showed that there is no "tension" between the horizontal and vertical determination of mental events by their neural supervenience bases and by prior mental events. As a consequence, the causal efficacy of mental states or events is no problem in the first place. The scheme of contextual emergence provides a constructive way to resuscitate mental causation.

Horizontal and vertical determination, suitably understood and combined, do not compete but complement one another. This result agrees with Yablo's approach based on the notion of proportionate causation and at the same time offers insight into theoretical and empirical details unavailable in his conceptual framework. Our work demonstrates how combining philosophical arguments with novel results in cognitive neurodynamics improves and refines our knowledge about mental causation.

Appendix: Generating Partitions and Symbolic Dynamics

Consider a partition $\mathcal{P} = (A_1, A_2, ..., A_m)$ over a state space X in which the states of a system are represented. Then a simple version of the *entropy* of the system is the well-known Shannon entropy

$$H(\mathcal{P}) = -\sum_{i=1}^{m} \mu(A_i) \; \log \, \mu(A_i) \, ,$$

where $\mu(A_i)$ is the probability that the system state resides in partition cell A_i.

The *dynamical entropy* of a system in a state space representation requires to consider its dynamics $\Phi : X \to X$ with respect to a partition $\mathcal{P}$ (Cornfeld et al. 1982):

$$H(\Phi, \mathcal{P}) = \lim_{n\to\infty} \frac{1}{n} H(\mathcal{P} \vee \Phi\mathcal{P} \vee ... \vee \Phi^{n-1}\mathcal{P})$$

In words, this is the limit of the entropy of the union of partitions of increasing dynamical refinement. The refinement is dynamical because it is generated by the dynamics Φ itself, expressed by $\Phi\mathcal{P}$, $\Phi^2\mathcal{P}$, and so forth.

A special case of a dynamical entropy of the system with dynamics Φ is the *Kolmogorov-Sinai entropy* Kolmogorov 1958; Sinai 1959

$$H_{KS} = \sup_{\mathcal{P}} H(\Phi, \mathcal{P}) \, .$$

This supremum over all partitions $\mathcal{P}$ is assumed if $\mathcal{P}$ is a *generating partition*, otherwise $H(\Phi, \mathcal{P}) < H_{KS}$. (Every Markov partition is generating , but not *vice versa*.) $\mathcal{P}_g$ minimizes correlations among partition cells A_i, so that they are *stable* under Φ and only correlations due to Φ itself contribute to $H(\Phi, \mathcal{P}_g)$. Boundaries of A_i are (approximately) mapped onto one another. Spurious correlations due to blurring cells are excluded, so that the dynamical entropy takes on its supremum.

Since the cells of a generating partition are dynamically stable, they can be used to define dynamically stable symbolic states, whose sequence provides a *symbolic dynamics* Γ (Lind and Marcus 1995). This dynamics is a *faithful representation* of the underlying dynamics only for generating partitions. The technical term "faithful" expresses that the underlying dynamics Φ and the properly constructed symbolic dynamis Γ are *topologically equivalent*.

Another way to say that Φ and Γ are topologically equivalent derives from the mapping π of states in X to symbolic states. If π is continuous and invertible, and its inverse π^{-1} is also continuous, π is called an intertwiner and we can write

$$\Gamma = \pi \circ \Phi \circ \pi^{-1} \, ,$$

such that the diagram in Sec. 3.2 is commutative: Starting with p, $\Gamma(m)$ can be equivalently reached via m or via $\Phi(p)$. The intertwiner π is topology-preserving if the partition yielding the equivalence classes of individual states in X is generating. (The topology of one state space is preserved in another one, if and only if any state change in one state space implies a state change in the other.) For generating partitions of X, there is therefore a one-to-one correspondence between statistical states in X and individual symbolic states.

The notion of topological equivalence is weaker than topological conjugacy, where the one-to-one correspondence applies to individual trajectories which can be parametrized pointwise in time. By contrast, statistical neural states have no individual trajectories but sets of trajectories, so that π cannot map statistical neural states onto symbolic mental states together with a one-to-one mapping of their time parameter. Therefore, we conjecture that this is another reason (in addition to those given in Sec. 5.1) why the one-to-one correspondence due to topological equivalence is not strong enough to be interpreted as an identity relation.

Acknowledgments

We are grateful to two reviewers of this journal for their helpful advice. Their questions and remarks were instrumental for clarifying particular issues and improving our argumentation at a number of places.

References

Adam, C. and P. Tannery (1904). Oeuvres de Descartes, 12 vols. Paris: J. Vrin.

Allefeld, C., H. Atmanspacher, and J. Wackermann (2009). Mental states as macrostates emerging from brain electrical dynamics. Chaos: An Interdisciplinary Journal of Nonlinear Science 19, 015102.

Allefeld, C. and S. Bialonski (2007). Detecting synchronization clusters in multivariate time series via coarse-graining of Markov chains. Physical Review E 76(6), 66207.

Amari, S-I. (1974) A method of statistical neurodynamics. Kybernetik 14, 201–215.

Amari, S-I., Yoshida, K., and Kanatani, K-I. (1977) A mathematical foundation for statistical neurodynamics. SIAM Journal of Applied Mathematics 33(1), 95–126.

Atmanspacher, H. (2007). Contextual emergence from physics to cognitive neuroscience. Journal of Consciousness Studies, 14 1(2), 18–36.

Atmanspacher, H. (2009). Contextual emergence. Scholarpedia 4(3), 7997.

Atmanspacher, H. (2012). Dual-aspect monism à la Pauli and Jung. Journal of Consciousness Studies, in press.

Atmanspacher, H. and P. beim Graben (2007). Contextual emergence of mental states from neurodynamics. Chaos and Complexity Letter 2, 151–168.

Baumgartner, M. (2008). Regularity theories reassessed. Philosophia 36(3), 327–354.

Bechtel, W. and J. Mundale (1999). Multiple realizability revisited: Linking cognitive and neural states. Philosophy of Science 66(2), 175–207.

Bennett, K. (2003). Why the exclusion problem seems intractable, and how, just maybe, to tract it. Noûs 37(3), 471–497.

Bennett, K. (2008). Exclusion again. In J. Hohwy and J. Kallestrup (Eds.), Being Reduced: New Essays on Reduction, Explanation, and Causation, pp. 280–305. New York: Oxford University Press.

Bishop, R. (2005). Patching physics and chemistry together. Philosophy of Science 72(5), 710–722.

Bishop, R. (2006). The hidden premises in the causal argument for physicalism. Analysis 66(1), 44–52.

Bishop, R. (2008). Downward causation in fluid convection. Synthese 160(2), 229–248.

Bishop, R. and H. Atmanspacher (2006). Contextual emergence in the description of properties. Foundations of Physics 36(12), 1753–1777.

Bishop, R. and H. Atmanspacher (2011). Causal closure of physics and free will. In R. Kane (ed.), The Oxford Handbook of Free Will, pp. 101–114. Oxford: University Press.

Block, N. (1997). Anti-reductionism slaps back. Noûs 31, Supplement: Philosophical Perspectives 11, Mind, Causation, and World 31(Supplement: Philosophical Perspectives, 11, Mind, Causation, and World), 107–132.

Block, N. (2003). Do causal powers drain away? Philosophy and Phenomenological Research 67(1), 133–150.

Bontly, T. (2002). The supervenience argument generalizes. Philosophical Studies 109(1), 75–96.

Broad, C. (1925). The mind and its place in nature. London: Kegan Paul, Trench, Turbne & Co.

Campbell, K. (1970). Body and mind. New York: Doubleday Anchor Books.

Chalmers, D. (2000). What is a neural correlate of consciousness? In T. Metzinger (Ed.), Neural correlates of consciousness: Empirical and conceptual questions, pp. 17–39. Cambridge, MA: MIT Press.

Cornfeld, I., S. Fomin, and Â. Sinaj (1982). Ergodic theory. Berlin/Heidelberg/New York: Springer.

Crisp, T. and T. Warfield (2001). Kim's master argument. Noûs 35(2), 304–316.

Dardis, A. (2008). Mental causation: the mind-body problem. Columbia University Press.

Davidson, D. (2001). Essays on actions and events. New York: Oxford University Press.

Elder, C. (2001). Mental causation versus physical causation: No contest. Philosophical and Phenomenological Research 62(1), 111–127.

Feigl, H. (1967/1958). The "mental" and the "physical": The Essay and a Postscript. Minneapolis: University of Minnesota Press.

Fell, J. (2004). Identifying neural correlates of consciousness: The state space approach. Consciousness and Cognition 13(4), 709–729.

Fodor, J. (1974). Special sciences (or: the disunity of science as a working hypothesis). Synthese 28(2), 97–115.

Fodor, J. (1997). Special sciences: Still autonomous after all these years. Noûs 31, Supplement: Philosophical Perspectives 11, Mind, Causation, and World 31(Supplement: Philosophical Perspectives, 11, Mind, Causation, and World), 149–163.

Gibbons, J. (2006). Mental causation without downward causation. The Philosophical Review 115(1), 79–103.

Goldman, A. (1969). The compatibility of mechanism and purpose. The Philosophical Review 78(4), 468–482.

Graben, P.b., Barrett, A., and Atmanspacher, H. (2009). Stability criteria for the contextual emergence of macrostates in neural networks. Network: Computation in Neural Systems 20, 178–196.

Haag R., Kastler D., and Trych-Pohlmeyer E.B. (1974) Stability and equilibrium states. Communications in Mathematical Physics 38, 173–193.

Harbecke, J. (2008). Mental Causation. Investigating the Mind's Powers in a Natural World. Frankfurt a. M.: Ontos.

Harbecke, J. (2010). The problem of mental causation formalized. Mind and Matter 8(1), 63–91.

Honderich, T. (1982). The argument for anomalous monism. Analysis 42(1), 59–64.

Huxley, T. H. (1893). On the hypothesis that animals are automata, and its history. In T. H. Huxley (Ed.), Collected Essays: Volume I, Method and Results,, pp. 195–250. London: Macmillan.

Kallestrup, J. (2006). The Causal Exclusion Argument. Philosophical Studies 131(2), 459–485.

Kant, I. (1902-1910). Gesammelte Schriften. Berlin: Reimer.

Kim, J. (1973). Causation, nomic subsumption, and the concept of event. The Journal of Philosophy 70(8), 217–236.

Kim, J. (1979). Causality, identity, and supervenience in the mind-body problem. Midwest Studies in Philosophy 4(1), 31–49.

Kim, J. (1984). Concepts of supervenience. Philosophy and Phenomenological Research 45(2), 153–176.

Kim, J. (1985). Supervenience, determination, and reduction. The Journal of Philosophy 82(11), 616–618.

Kim, J. (1997). The mind-body problem: taking stock after forty years. Nous 31, 185–207.

Kim, J. (1998). Mind in a physical world: An essay on the mind-body problem and mental causation. Cambridge: MIT Press.

Kim, J. (2003). Blocking causal drainage and other maintenance chores with mental causation. Philosophy and Phenomenological Research 67(1), 151–176.

Kim, J. (2005). Physicalism, or something near enough. Princeton: Princeton University Press.

Kolmogorov, A.N. (1958). New metric invariant of transitive dynamical systems and endomorphisms of Lebesgue spaces. Doklady of Russian Academy of Sciences 119, 861–864.

Leibniz, G. W. (1875-1890). Sämtliche Schriften und Briefe. Berlin: Akademie-Verlag.

LePore, E. and B. Loewer (1989). More on Making Mind Matter in Philosophy of Mind. Philosophical Topics 17(1), 175–191.

Lind, D. and B. Marcus (1995). An introduction to symbolic dynamics and coding. Cambridge: Cambridge Univ Press.

Loewer, B. (2007). Mental causation, or something near enough. In B. McLaughlin and J. Cohen (Eds.), Contemporary Debates in Philosophy of Mind, pp. 243–264. Blackwell Publishers.

MacDonald, C. and G. MacDonald (1986). Mental causes and explanation of action. The Philosophical Quarterly 36(143), 145–158.

Malcolm, N. (1968). The conceivability of mechanism. The Philosophical Review 77(1), 45–72.

Marras, A. (2007). Kim's supervenience argument and nonreductive physicalism. Erkenntnis 66(3), 305–327.

McLaughlin, B. (1992). The rise and fall of British emergentism. In A. Beckermann, H. Flohr, and J. Kim (Eds.), Emergence or reduction?: essays on the prospects of nonreductive physicalism, pp. 49–93. Berlin/New York: Walter de Gruyter.

McLaughlin, B. and K. Bennett (2008). Supervenience. In E. N. Zalta (Ed.), The Stanford Encyclopedia of Philosophy (Fall 2008 ed.).

Montero, B. (2003). Varieties of causal closure. In S. Walter and H. Heckmann (Eds.), Physicalism and mental causation: the metaphysics of mind and action, pp. 173–187. Exeter: Imprint Academic.

Nagel, E. (1961). The Structure of Science: Problems in the Logic of scientific Explanation London: Routledge and Kegan Paul.

Noordhof, P. (1999). Micro-based properties and the supervenience argument: A response to kim. Proceedings of the Aristotelian Society 99, 109–114.

Papineau, D. (2002). Thinking about Consciousness. New York: Oxford University Press.

Place, U. T. (1956). Is consciousness a brain process? British Journal of Psychology 47, 44–50.

Primas, H. (1998). Emergence in exact natural sciences. Acta Polytechnica Scandinavica 91(1998), 5–83.

Primas, H. (2007). Non-Boolean descriptions for mind-matter problems. Mind and Matter 5(1), 7–44.

Putnam, H. (1975). Philosophy and our mental life. In H. Putnam (Ed.), Mind, Language and Reality: Philosophical Papers, vol. 2, pp. 291–303. Cambridge: Cambridge University Press.

Putnam, H. (1980/1967). The nature of mental states. In N. Block (Ed.), Readings in philosophy of psychology. Vol. 1, pp. 223–231. Cambridge: Harvard University Press.

Robb, D. and J. Heil (2009). Mental causation. In E. N. Zalta (Ed.), The Stanford Encyclopedia of Philosophy (Summer 2009 ed.).

Ryle, G. (1949). The concept of mind. London: Hutchinson.

Sachse, C. (2007). Reductionism in the Philosophy of Science. Frankfurt am Main: Ontos.

Schiffer, S. (1987). Remnants of meaning. Cambridge, MA: Mit Press.

Schopenhauer, A. (1980). Sämtliche Werke. Darmstadt: Wissenschaftliche Buchgesellschaft.

Schröder, J. (2007). Mental Causation and the Supervenience Argument. Erkenntnis 67(2), 221–237.

Shoemaker, S. (2001). Realization and mental causation. In C. Gillett and B. Loewer (Eds.), Physicalism and its Discontents, pp. 74–98. Cambridge/New York: Cambridge University Press.

Sinai, Ya.G. (1959). On the notion of entropy of a dynamical system. Doklady of Russian Academy of Sciences 124, 768–771.

Smart, J. (1959). Sensations and brain processes. The Philosophical Review 68(2), 141–156.

Sosa, E. (1984). Mind-body interaction and supervenient causation. Midwest Studies in Philosophy 9(1), 271–281.

Spinoza, B. d. (1999 (1677)). Ethica Ordine Geometrico demonstrata et in quinque Partes distincta. Hamburg: Meiner.

Sturgeon, S. (1998). Physicalism and overdetermination. Mind 107(426), 411.

Takesaki, M. (1970). Disjointness of the KMS states of different temperatures. *Communications in Mathematical Physics* 17, 33–41.

Thomasson, A. (1998). A nonreductivist solution to mental causation. Philosophical Studies 89(2), 181–195.

Yablo, S. (1992). Mental causation. The Philosophical Review 101(2), 245–280.