

Pushing brains: Can Cognitive Neuroscience Provide Experimental Evidence for Brain-Mind Causation?

Abstract

What makes the issue of causal relations between mental and cerebral events so special? And is there experimental evidence from neuroscience for this sort of causation? To answer these questions, the issue of brain-mind causation is considered against the background of the mind-brain problem and the theory of causation in general. Then, one empirical study from cognitive neuroscience is discussed as an example of how the correlations of mental and cerebral events and processes are investigated in current research. From the prevailing empirical studies, it is obvious that neuroscience can only demonstrate concurrence of cerebral and mental events, not an additional causal relationship between them. The decision for a certain causal interpretation is based on precedent commitments to particular philosophical attitudes concerning, among other things, the mind-brain problem, scientific explanation, and the issue of reductionism. Finally, it will be argued that it may make sense to interpret brain-mind relations causally, provided that we do not understand causation in terms of some sort of physical connection. Current process theories are not applicable to the case of brain-mind causation, because they are largely conceptualized in terms of the physical, so that there is no conception of how they could be applied to an instance of causation that involves a non-physical relatum.

Martin Kurthen¹

*Schweizerisches Epilepsie-Zentrum
Zürich, Switzerland*

Causality is the presumably necessary relationship that holds between two repeatedly occurring contiguous events,² if the tokens of one of them (the effect) are the consequences of, that is, consistently follow, the tokens of the other (the cause). Causation is the act or the process or performance by which this relationship is established.

The present paper provides some remarks about what makes the issue of causality between mental and cerebral events so special and whether experimental evidence from neuroscience exists for this sort of causation. I will talk of *mind-brain causation* for the case when mental events are conceived as causes, while cerebral events are conceived as effects

¹ Schweizerisches Epilepsie-Zentrum, Bleulerstrasse 60, 8008 Zürich, Switzerland
Tel. 0041-(0)44-3876111; Fax 0041-(0)44-3876397; martin.kurthen@swissepil.ch

² The question of whether or not this relationship has to be construed as a necessary one has been one of the major issues in the philosophy of causation. However, it is disputable if it still makes sense to speak of "causation," if there is only a contingent relationship between two events. Perhaps less theory-laden terms like "regularity of succession" would be more suitable in that case. To start with, I shall view causality as a necessary relation.

and of *brain-mind causation* for the reverse case, when cerebral events are causes, while mental events are effects.³

In what follows, I will not discuss the special problem of mind-mind causation, namely, the question of whether mental events can be causes (and effects) of other mental events, provided that these events are not identical with cerebral events. Nor shall I cover the topic of brain-brain causation, assuming that this issue collapses into the general discussion of physical causation altogether. This is because two cerebral events are just two physical events, and causation is conceptualized as applying primarily to the physical realm. Admittedly, there may be special aspects of causation when we consider biological or physiological events as subclasses of physical events, in contrast with events that are current objects of physics. For example, Walter J. Freeman (1999) has argued that brain-mind causation is special in that it has a circular structure (it is organized in cycles or spirals of self-organizing microscopic and macroscopic brain states) rather than linear (conceived as a causal chain of cerebral events between stimulus and response). For him, awareness and intentionality arise as higher-order phenomena within these dynamic cycles: “Intentional acts are produced by the self-organized microscopic neural activity of cortical and subcortical components in the brain. Awareness supervenes as a macroscopic ordering state, that defers action until the self-organizing microscopic process has reached closure in reflective prediction” (Freeman 1999, p. 143). Freeman supplements this neurophysiological interpretation with a general attitude towards causation that mixes elements of nominalism and constructivism. But although the introduction of self-organization, circularity, and environmental embedding into neurophysiology marks a major progress in our understanding of higher brain functions, Freeman’s approach does not touch the question of brain-mind causation as discussed in the present paper. Somewhat misleadingly, he speaks of “circular causality,” where something like “circular causation” seems to be meant: Causation runs in circles, but causality as such is not “circular.” Circular brain processes can still be treated as a particular form of neural correlates of the mental phenomena that “supervene” on or are “produced” by them, so that the question of causal relations between these cerebral and mental tokens remains. What Freeman describes is a special form of brain-brain causation, associated with supervening or emerging mental phenomena.

In the following two sections, the question of brain-mind causation will be considered against the background of the mind-brain problem and the theory of causation in general. In the fourth section, I will present and discuss one empirical study from cognitive neuroscience as an example of how the correlations of mental and cerebral events and processes are investigated in current research. In section (5), some general remarks on the general scope of experimental neuroscientific evidence with respect to neurophilosophical questions will follow. Finally, I will maintain that the choice of a causal vs. non-causal interpretation of brain-mind correlations cannot be grounded in neuroscientific evidence. However, I will suggest that it may still make sense to interpret brain-mind relations causally, provided that we do not understand causation in terms of some sort of physical connection. It will be

³ Hence, I will not cover the problem of “mental causation” proper, which applies to causality between mental and behavioral events. The persuasiveness of realism about mental causation is largely due to the fact that mental-behavioral causation is an essential part of our commonsensical picture of ourselves as human beings or rational agents. If we try to argue for mental-cerebral causation, however, we will have to get along largely without that intuitive plausibility. While it seems obvious that my intention to raise my arm causes my arm to raise, it is not part of our commonsensical understanding (unless we are neuroscientists) that my intention to raise my arm causes my brain to build up a certain pattern of activation in the primary motor area of the cerebral hemisphere contralateral to the arm to be raised. But this is exactly what would have to be the primary effect of my intention, if it should also manage to cause my arm to raise. To keep the text shorter, I will sometimes use brain-mind causation for the whole complex of brain-mind and mind-brain causation.

concluded that although there cannot be specific evidence for brain-mind causation, we may nonetheless find clues for brain-mind causation: Given that there is causation of a certain type (for any pair of relata, physical or non-physical), then there can be evidence for that sort of causation between cerebral and mental events, not only between cerebral events and other physical events.

Background

The Mind-Brain Problem

Under certain theoretical conditions, the problem of brain-mind causation itself collapses into the problem of brain-brain causation. For example, if you decide to become an identity theorist (claiming that mental events are type-identical with cerebral events), you can treat the problem of brain-mind causation as part of the problem of brain-brain causation. This illustrates that the problem of brain-mind causation comes in different ways dependent on prior decisions concerning more general philosophical questions. Particularly, the choice of a determined attitude towards the mind-brain problem (understood as the general question of the how to specify the relation between mental and cerebral events or processes⁴) will eminently influence the guise in which the problem of brain-mind causation can subsequently appear.

Any reductive materialist can treat brain-mind causation as a special case of brain-brain causation, because for him, the mental reduces to the cerebral (see above). A parallelist will not have to bother with brain-mind causation because his position is defined by the absence of this sort of interaction. An epiphenomenalist, that is, a dualist who claims that mental events are caused by physical events without being causally efficacious themselves, will have to specify brain-mind causation, but not mind-brain causation, which for him is excluded by definition. Partial or complete deliverance from worries about causation is, however, not without costs, since reductive materialism, epiphenomenalism and, most of all, parallelism display major problems of internal plausibility. (How could a mental event with all its experiential aspects be identical or reduce to a cerebral event?⁵ How can epiphenomenalism be saved from collapsing into reductionism? What could be the mechanism by which mental and cerebral processes are held in parallel lines, etc.?)⁶ The position that most strongly evokes the problem of brain-mind causation is dualist interactionism, according to which the mental and the cerebral are substantially different, yet interact. By construction, this position is in need of specifying brain-mind causation in both directions. Consequently, one of the main objections to this position has been that it is hard to even conceive how tokens of two different substances could interact causally at all.⁷

⁴ See Kim (1996) for an overview.

⁵ In many respects, the experiential or phenomenal aspects (the *qualia*) of a mental event are the touchstone for physicalistic theories. This is why conscious mental events are usually chosen as examples, when the difficulties of mind-brain physicalism are discussed.

⁶ Cave: Many positions have initially been construed to solve the *mind-body* problem, not the *mind-brain* problem. The problem of causal interaction, too, has often been formulated in terms of *mind-body* causation, not *mind-brain* causation. The pros and cons of any of these positions may differ depending on the broader context of *mind-body* versus *mind-brain* relations (see footnote 1 above for mental causation).

⁷ To be fair, the problem of how a causal interaction literally takes place is not resolved if the relata are just allocated to the same substance. Quite the contrary, for any causal process, the exact mechanism of causation seems to remain ineffable (see below, section 6). The possibility remains, however, that the determination of such a mechanism will only succeed for causation within substances (whatever that means).

Furthermore, dualist interactionism seems to be incompatible with the broadly naturalistic worldview that actually dominates the philosophy of mind and, of course, neuroscience. Today, interactionist dualism has largely fallen from grace, a fact that I will take as an ostensible justification to pass over this position in what follows.⁸ Instead, I propose to tentatively proceed with non-reductive physicalism (see the following paragraph), a position that is acceptable both for many philosophers of mind and for neuroscientists, representing a compromise between the intuition that the mental is not reducible to the physical, and the belief that our ultimate (scientific) explanation of the world will only deal with physical entities. Since in the present paper, it is intended to discuss the possibility of neuroscientific evidence of brain-mind-causation, it would not make sense to start from a philosophical position that either permits only brain-brain causation in the first place (like reductive materialism) or is in general conflict with the worldview of neuroscience (like substance dualism).

Today many philosophers of mind favor some sort of non-reductive physicalism (or materialism), claiming that ultimately “all there is” are physical or material entities (or the entities invoked by physics),⁹ although the mental, while asymmetrically dependent on the physical, is not reducible to these entities. For example, mental properties may be resistant to intertheoretical reduction owing to their multiple realizability (see footnote 10).¹⁰ In metaphysical terms, the position that mental phenomena are (irreducible) non-physical properties of physical phenomena has been characterized as *property dualism* (Kim, 1996). For non-reductive physicalists, the question of brain-mind causation arises again. Today, the most thoroughly discussed version of non-reductive physicalism is the supervenience thesis, according to which the mental supervenes on the cerebral, in that there can be no change in mental properties without an underlying change in cerebral properties (but not vice versa). Although this may look like an attractive compromise, it has been controversial whether the supervenience relation is strong enough for physicalists to accept or whether it allows unacceptable violations of physicalist intuitions (see McLaughlin & Bennett, 2005 for a review). There is as yet no general consensus on how to formulate a non-reductive physicalist theory of mind and brain that does not impend to fall back into either reductionism (or epiphenomenalism) or dualism. The basic, though vague, idea of non-reductive physicalism is that against the general background of a physicalist ontology, the mental is asymmetrically dependent on the cerebral, without being reducible to it. One way of expressing this idea is to

⁸ This is not to say that interactionism has been abandoned completely. There is one tradition according to which mind-brain (causal) interaction could occur on a quantum level (e.g. Stapp, 2005). Other philosophers have argued that standard arguments against substance dualism have failed, contrary to widespread opinion (Lowe, 2000, 2006; Mills, 1997).

⁹ Unfortunately, the ontological business is not finished by declaring that I am not a dualist, but a materialist or physicalist. If all there are are material entities, then what exactly is “matter”? If physics prescribes ontology, then I am challenged to decide whether I refer to current physics (which might turn out to be wrong) or to an ultimate future version of physics (the entities of which we cannot anticipate at present). An alternative is “naturalism,” the general claim that “nature exhausts reality” (Papineau, 1993) (there are no “supernatural” entities) and that the natural sciences provide the best (or even the only) way to explain this reality (but again, the entities invoked by the ultimate natural sciences are currently unknown, and the present natural sciences may turn out to have misconstrued their ontology).

¹⁰ In the context of the mind-brain problem, *reduction* is often understood as intertheoretical reduction, that is, the reduction of mentalistic theories to physicalistic theories, or conceptual reduction, that is, the analysis of mental predicates in terms of physical predicates (as contrasted with the stronger *ontological reduction* of mental entities to physical entities, as postulated by the mind-brain identity theory).

say that mental events are “realized” by cerebral events (but not vice versa). *Realization* is understood as an asymmetrical and synchronic (and co-locational, see Jaworski, 2006, p. 288) dependence relation (in contrast to the causation relation, which is asymmetrical and diachronic).¹¹ This sort of “realization physicalism” (Polger, 2007, p. 233), however, bypasses rather than solves the causation debate in that it avoids the demand of considering brain-mind causation directly. The realization physicalist can hold that certain brain events cause other brain events to occur, and that these other brain events are the realizers of certain mental events. Anyway, within the field of non-reductive physicalism, the problem of causation has mainly been discussed in terms of traditional mental causation: How can a token of behavior be (additionally) caused by a supervenient mental event, if that behavior is already sufficiently caused by the underlying cerebral event? How can my intention to raise my arm cause my arm to rise, if that behavior is already caused by a firing pattern in the contralateral motor cortex of my brain? This is the problem of causal overdetermination (Funkhouser, 2002) that any theory of mental causation, based on a not strictly monistic mind-brain philosophy, has to face. The potential solution that it is the mental event, not the cerebral event, that causes the token of behavior, stands against the widely accepted belief that the principle of (causal) completeness of physics also holds in neuroscience. There is some controversy about how to formulate that principle (Lowe, 2000), but many would probably agree on something like “All physical events have complete physical causal histories” (Marcus, 2005, p. 28). Some philosophers have argued that acceptance of the principle of completeness still leaves room for mental causation in a non-reductive approach (Lowe, 2000, 2005; MacDonald, 2007; Mills, 1997). The principle of completeness is distinguished from the stronger thesis of causal closure, which means that “physical events cannot interact causally with non-physical events” (ibid.) or, in other words, that nothing non-physical can causally affect the physical. According to the principle of completeness, any cerebral event with a cause will have other cerebral (or at least physical) events as its causes. According to the principle of closure, no mental (non-physical) event can causally affect cerebral events. As for mental causation, this seems to imply that mental events cannot be causes of behavior, because behavioral events are caused by cerebral events, and causality is (widely held to be) a transitive relation. One way to reconcile physicalism with realism concerning mental causation is to attribute different types of causation to cerebral and mental events (see, for example, Dretske’s [1988] concept of triggering vs. structuring causes of behavior).

As said above, in the present text I will not deal with behavior, but only with brain-mind-relations: Will there be brain-mind causation if the mental “depends” on the cerebral without being reducible to it?

¹¹ There is some controversy of how best to specify the realization relation. Realization was introduced in the context of machine functionalism by Putnam: Realization is the relation that holds between machines and their programs and, by analogy, between brains and minds (Putnam, 1975). The program is not “identical” to physical states of the machine, but it is implemented by them, instantiated in them, or realized by them. Following the general concept of functionalism, realization can be understood as role-occupancy: A cerebral token realizes a mental token if it occupies the very role (or, in other words, has the very function, see Polger 2007, p. 251) that characterizes the mental token (McLaughlin, 2007). If a mental property is taken to be a higher-order property (namely, the property of having some property or other that occupies a certain role), the nature of the first-order property of occupying the role is left open. If the role is occupied by a physical property, then this physical property can be said to be the realizer of the mental property. If there is more than one first-order property to occupy the role, then the mental property can be said to be multiply realizable. In that case, the higher-order property is not identical to the first-order property. This is compatible with property dualism as a version of non-reductive physicalism: Although mental properties may (contingently) always be realized by physical properties, they are not identical with or reducible to them (see McLaughlin, 2007, p. 152).

General approaches to causation

The notion of asymmetrical mind-brain dependency, as part of the doctrine of non-reductive physicalism, seems to rule out the possibility of mind-brain causation, even if the mental is conceived as “irreducible.”¹² I will take this as a welcome, although again ostensible, reason to confine myself to the issue of brain-mind causation for the rest of this paper and skip the issue of mind-brain causation for the present purposes. What is left, then, is the question of whether the neural correlate of a mental event¹³ can be said to “cause” that event and whether there is or can be neuroscientific evidence for such a causation.¹⁴

Is it conceivable that a mental token is literally caused by its neural correlate? This is at least partly dependent on prior decisions about how causality in general is to be conceptualized. Unfortunately, the issue of causation and causality, independent of the special problem of brain-mind causation, has itself remained unresolved in philosophical and scientific discussions. The choice of a certain theory of causation, however, will heavily influence our subsequent approach to brain-mind and mind-brain causation. Imagine, for example, that we would decide to be causal eliminativists claiming that there is no such relation like causality in the world. This would free us from the whole problem of brain-mind causation.¹⁵ To be less radical, if we accept that there is causation, we may understand causal processes and their mechanisms in very different ways, and we may thus pursue different approaches to the metaphysics of causality. The first philosophical decision to be made in that context concerns the adequate placement of a theory of causality within the philosophical disciplines, namely,

- Ontology: Here, the main approaches are that of realism (causality is an objective feature of the world) versus nominalism/instrumentalism (causality is just a concept, but a useful one for explanation) or even eliminativism (there is no such thing as causality).

¹² This (if accepted) does not seem to deter non-reductive physicalists from being realists about mental causation. It seems that as a realist, one has to either abandon the thesis of the transitivity of causation, or hold that mind-body causation differs from mind-brain and brain-body causation in significant respects (see above, footnote 5).

¹³ What exactly is meant by the “neural correlate” of a mental event? This question has been discussed extensively for the case of the *neural correlate of consciousness* (NCC) in the philosophy of mind (Metzinger, 2000). To put it simply, an NCC can be defined by there being a mapping from states of a neural system to states of consciousness, where the respective state of the neural system is sufficient for the corresponding state of consciousness (Chalmers, 2000). Conceptual work on the NCC is important for the discussion of the possibility of the reduction of mental (or just conscious) states to neural states, but the issue of the NCC is largely neutral with respect to the problem of brain-mind causation as discussed in the present manuscript (see Fell, Elger & Kurthen, 2004). Even if it were uncontroversially true that there is an NCC in the above sense (but see Noe & Thompson, 2004, for a contra argument), this would leave us with various options for conceptualizing a causal relation between a mental event and its NCC.

¹⁴ If we want to take non-reductive physicalism as a background doctrine, we must accept that we will have to discuss this question in spite of a lack of general consensus about how to conceive of mental events that are (1) parts of a world where ultimately all there is are physical events, and yet (2) irreducible to physical events. Here, we would get involved in deeper philosophical issues: What is meant by reduction? What does it mean to “be there” at all? What exactly is meant by “physical”? Why cling to a metaphysics of events (rather than properties, facts, processes, or whatever?) in the first place? See the continuous text above for at least some remarks on these problems.

¹⁵ In that case, however, we would have to make the whole of philosophical theorizing and scientific explanation work without the concept of causality. Some philosophers think that for that endeavor, the prospects for success are even worse than for causation theory, so that we would do well to not eliminate causality.

- Epistemology: The two most famous epistemological positions are the Kantian of transcendental idealism (causality as an a priori category of the human *Verstand*), and the Humean regularity theory (causality as an associative concept derived from the observation of the regular succession of events).
- Conceptual analysis: In this approach, the goal is not to investigate causality as a feature of the world, but to analyze our current, ordinary-language based, everyday understanding of causality.
- Philosophy of science: In contrast to conceptual analysis, the endeavor is something like “empirical analysis,” an investigation and interpretation of the role of causality in the actual natural sciences.

The second decision to be made refers to the choice of a theoretical approach to causality. The numerous strategies on hand can be grouped dichotomically (Collins et al., 2004) into *physical connection* accounts on the one hand and *nomological entailment* accounts on the other.¹⁶ Physical connection accounts try to grasp causation by specifying some sort of “quantity” the “transfer” of which constitutes the causal relation (transfer of an energy, of a “conserved quantity,” of a trope...). By contrast, nomological entailment approaches generally take “*c* to be a cause of *e* just in case it follows, from the proposition that *c* occurs, together with the proposition that encapsulates the fundamental laws, that *e* occurs” (Collins et al., 2004, p. 14). If we accept a broad notion of nomological entailment, we can also subsume probabilistic theories of causation (*c* causes *e* if the occurrence of *c* raises the probability of *e*’s occurrence) and counterfactual theories of causation (basically: *c* is a cause of *e* if, had *c* not occurred, then *e* would not have occurred either) under this term (see Collins et al., 2004, p. 15). An even more basic distinction is that between reductionist and non-reductionist theories of causation, or as Schaffer (2008) says, reductionist and “inflationist.” The reductionist holds that “causal facts can be reduced to categorical plus nomological facts” (Collins et al., 2004, p. 12). Categorical properties involve what objects are actually like, while dispositional properties seem to “point beyond” the categorical in that they specify the potentiality of objects, the effects that might occur under certain conditions. For example, there is the effect of being ignited if a match is struck: The match has the “power to ignite” as a dispositional (or “modal”) property, in contrast to its categorical or “occurrent” properties like having a tip made of certain chemicals (see Schaffer, 2008). For the reductionist, causation reduces to the categorical plus the laws of nature, understood as the laws “that govern what happens” (Collins et al., 2004, p.12): There is nothing more than patterns of events plus the (strong but contingent) regularities these patterns display. For the non-reductionist (or inflationist), on the other hand, causation is a fundamental trait of the world, an additional feature or “power” that turns the connection between cause and effect into a necessary relation.¹⁷ All these theories have their respective strengths and weaknesses, which can not be weighed up against each other in the present paper (see again Schaffer, 2007, for a short overview).

¹⁶ Schaffer (2007) uses a similar dichotomy, namely that of “process” accounts where “causing is physical producing,” and “probability” accounts, where causing is “making more likely.” For him, the former comprise, among others, theories in which causation is construed as a flow of energy, or property transference, while the latter include analyses of causation in terms of nomological subsumption, statistical correlation, and counterfactual dependence. In a similar spirit, Fell et al. (2004) distinguish “efficient causality” from “explanatory causality.”

¹⁷ For the pros and cons of reductionism see Schaffer (2008), Esfeld (2007), Beebe (2006), and Chakravartty (2005).

The third philosophical decision deals with the determination of the causal relata, for example,

- concrete relata like events, features, or tropes, versus
- abstract relata like properties or facts.

The choice of the relata may depend on explanatory preferences. For example, only immanent relata like features seem to be able to enter into concrete interaction, which, after all, is what happens in causation.¹⁸ On the other hand, because in absence-related causation there are no immanent entities involved, the case that absences (non-occurrences) enter into causal relations can only be handled by a transcendent approach.¹⁹

All the different approaches have their advocates and opponents, and it is difficult to imagine that one singular account will eventually prevail. Actually, the extensive literature on this topic seems to suggest the conclusion that causation is not uniform: There may be different types of causation, requiring different approaches and perhaps different relata (Collins et al., 2004). This has led some researchers to opt for hybrid theories of causation, trying to combine the advantages of, for example, process and probability theories (Schaffer, 2001). For the present purposes, I will confine the analysis to mental and cerebral *events* (understood as particulars, as entities that occupy certain spatiotemporal regions) as relata, and I will assume that causality, if present, is a feature of the world, not just an a priori *Verstandeskategorie* or epistemological construction. Immanent relata seem to be suited for discussions about empirical neuroscience (where experimental data refer to immanent entities like neural activation patterns) more than transcendent relata. Without realism concerning causation, the consideration of neuroscience seems to lose its momentum right from the start (why bother with neuroscientific evidence for causation, if causality is nothing but a habit of human thought?). Should we then try to capture brain-mind causation in terms of connection (or process) or rather by means of a nomological entailment approach? To find an answer to this question, it may be helpful to first consider the actual practice in cognitive neuroscience research. In the following section, I will pick out and present as an example one empirical study which is paradigmatic or at least typical for our current research practice.

Cognitive Neuroscience: Mind-Brain Correlations

The typical research procedure in experimental cognitive neuroscience runs as follows: First the object of study is determined, usually by defining a mental subfunction or faculty of interest. Then, this mental function is operationalized for an experimental procedure. After that, the experiment is performed in healthy persons or in patients with well-defined cerebral dysfunctions. The empirical data are obtained about the cerebral processes or events that occur when subjects apply the cognitive function of interest in controlled circumstances. The data obtained are then analyzed statistically, and the results are interpreted against the background of prior knowledge about cerebral functions, processes, and events as “underlying” or “realizing” their mental counterparts. These interpretations are mostly outlined under the assumption of weak *localizationalism*, namely the belief that circumscribed

¹⁸ To this, the advocate of transcendent relata may answer that the concrete “causal pushing” can be done by substitute entities like objects, while facts remain the relata of causality (see Schaffer, 2007, for an overview of the pros and cons).

¹⁹ This leaves room for the counterargument that absences only seem to be causal, or that they can be construed as immanent (negative) properties (see again Schaffer, 2007, for an overview).

areas of the brain contribute asymmetrically to the realization of distinct mental processes and events. Out of the vast literature in the field of cognitive neuroscience, I will now pick out one study that is quite typical in its general approach and methods. In fact, this very study has been used by an interactionist dualist (Stapp, 2007) for an argument for the presence of causal interaction between mind and brain (see below).

Like many other researchers in cognitive neuroscience, Ochsner et al. (2002) used functional magnetic resonance imaging (fMRI) to investigate neural correlates of a certain cognitive function. They turned their attention to “reappraisal,” which in the context of that study refers to the cognitive reappraisal, the “rethinking” of emotionally aversive events in nonemotional terms (with the goal of reducing negative affect). According to prior knowledge, they hypothesized that this reappraisal requires (1) strategies for cognitively “reframing” an emotional event (for this, the lateral prefrontal cortex would be the candidate brain region according to prior evidence), (2) monitoring interference between top-down reappraisal and bottom-up emotional evaluation (candidate region: dorsal anterior cingulate cortex), and (3) ongoing reevaluation of the relationship between external stimuli and internal states (candidate region: dorsal aspects of the medial prefrontal cortex). Reappraisal was operationalized for an experimental setup, where standardized color photos with negative emotional load (control condition: emotionally neutral color photos) were presented to normal participants who were instructed to either passively attend to the feelings elicited by the photos (“attend” condition) or actively reinterpret them to eliminate the negative emotional response (“reappraise” condition). Reappraisal was trained systematically before the experiment started. fMRI data were then acquired from the participants during a total of 114 trials with the two stimulus classes and instructional conditions. Adequate statistical processing yielded activation maps (more precisely, maps of regional differences in BOLD effects, that is, differences in oxygen utilization of arterial blood perfusing the respective regions) for the respective stimulus classes and test conditions. As reappraisal-sensitive regions (defined by a greater activation on “reappraise” than on “attend” trials for negatively loaded photos), the left lateral prefrontal cortex, the dorsal medial prefrontal cortex, and additional areas (left temporal pole, right supramarginal gyrus, left lateral occipital cortex) were identified. The main emotion-sensitive region (greater activation on “attend” trials) was the medial orbitofrontal cortex. Furthermore, during effective reappraisal there was decreased activation in the medial orbitofrontal cortex and the amygdala, correlated with increases in the lateral prefrontal cortex. In their discussion of these results, the authors state that their reappraisal-sensitive regions largely overlap with those regions that are known for their role in working memory and response selection. They take their results as support for the hypothesis that the lateral prefrontal cortex contributes to the modulation of emotional processing.

What can a study of this type tell us about brain-mind causation? If we take a look at the general conceptual architecture of Ochsner’s experiment (and, basically, of virtually all the experiments in cognitive neuroscience, with some minor deviations, see figure 1), we first see that what is actually registered in this experiment is not mental and cerebral events (or processes, considering the process character of reappraising and attending), but rather tokens of behavior (participants watching single photos for some seconds) and physiological changes (BOLD effect) that are assumed to differentially reflect brain activity. On the upper level of publicly observable events in figure 1, we also have the test instruction and the application of the stimuli (visual presentation of photos). The cognitive function of interest (reappraisal) is embedded in the experimental setup plus the test instruction. Setup and instruction carry a “cognitive load”: They are so construed that the participants should use the function in question while running through the test procedure. The design of the setup and the instruction

are based on a certain pre-existing theoretical conception of the cognitive function. It is assumed that distinct mental events or processes (tokens of reappraisal or its subprocesses) occur when the participants behave according to the test instructions. Hence, mental events are not directly measured in this experiment; rather, their occurrence is inferred according to the theoretical load of the test instruction and procedure. As far as possible, the occurrence of the respective tokens of reappraisal mentations are retrospectively verified in terms of interviews (in the Ochsner et al. study, participants had to report their reappraisal strategies after the trials). Cerebral events are not measured either: All we have is a pattern of behavior (watching photos) and a correlated physiological change (BOLD effect). Cerebral events (neural correlates of reappraisal), too, are inferred, according to the established interpretation of the BOLD effect (regionally increased oxygen utilization as a marker of increased neural activity in the time window of the behavioral response and the assumed concurrent mentations). Hence, both mental and cerebral events enter only indirectly into the whole procedure. As for the timing of the measurements, physiological changes are usually recorded in a time window from the beginning to the end of stimulus presentation (time of presentation of each photo), and the time course of the changes within that window is not taken into account. Concerning the reappraisal mentation, one can assume that it starts no earlier than around 100 milliseconds after stimulus onset (this is the time usually needed for a visual stimulus to become represented in the visual cortex) and end, approximately, with the end of stimulus presentation; again, the time course within that window is not considered. For our consideration of brain-mind causation, we thus find:

1. In an experiment of the type presented, there is no direct data from mental and cerebral events. Even if we would chose a method that allowed for a more direct measurement of cerebral events (for example, recording of action potentials from single neurons), it would still be impossible, at least with current technical means, to record the whole cerebral event that correlates to the respective reappraisal mentation. (This event will probably involve millions of single neurons, in changing constellations and variable activation patterns). Furthermore, a more direct measurement of the mental events is impossible, because only behavioral correlates of mental events (e.g., verbal reports) are publicly accessible, not these events themselves.
2. Due to both the experimental procedure and the indirect access to mental (and, to different degrees, also to cerebral) states, it is not possible to determine the exact temporal relations between the involved cerebral and mental events. Furthermore, there are good reasons to assume that the mental or cognitive events in particular are temporally “blurred” concerning their beginning and their end. After all, reappraisal, whatever its subprocesses, cannot be expected to start or end abruptly from one millisecond to another. Rather, it is plausible to assume that it gradually commences within some milliseconds. That these events are temporally blurred may already be a consequence of their not being unitary processes, but rather theoretical idealizations or psychological constructions. Lastly, there seems to be no way to reach the fine-grained time resolution for the measurement of conscious mental states that can be achieved in the neural domain (Fell et al., 2004).
3. Finally, there are no neuroscientific data at all about the (non-temporal) relation(s) between the mental and cerebral events involved in the participants’ behavior. The only “relation” that can be obtained is concurrence in a vaguely determined time window. The measurement data are completely neutral and indifferent with respect to the particular relation that may subsist. They are absolutely compatible with there being a relation of supervenience, parallelism, or identity between cerebral and mental events, and they do not allow to empirically differentiate

a relation of causation from that of mere concurrence. This lack of evidence concerning the brain-mind relation seems to be a matter of principle, since the neuroscientific data obtained in experiments of this type simply carry no information about the type of relation that may hold between the two classes of events. Would these conclusions have been different if we had considered experiments from other fields of neuroscientific research? After all, one might argue that fMRI studies as such are controversial in many methodical respects (Henson, 2005; Brown & Eyster, 2006), and the sometimes naive localizationism that prevails in this field of research (Knight, 2007) has provoked mocking phrases like that of “the new phrenology” (Uttal, 2003). Nonetheless, the general limitations of data from neuroscientific experiments hold independently of the particular shortcomings of fMRI research, at least as far as the issue of causation is concerned. Let us shortly consider two other, methodically less intricate neuroscientific approaches which seem to yield more “direct” data, namely, electrical stimulation mapping and lesion studies. In the presurgical evaluation of epilepsy, electrical stimulation mapping in humans is applied to delineate eloquent brain areas from the epileptogenic tissue to be resected. Low-amplitude electrical stimuli of short duration are directly applied to circumscribed areas of the cerebral cortex via subdural electrodes in order to elicit sensory or motor responses from eloquent cortices. In a famous study, Blanke et al. (2002) induced out-of-body experiences (OBE) by direct stimulation of the right angular gyrus of an epileptic person via a subdural grid electrode. Did not this experiment show that a direct stimulation of the brain may “cause” a mental state to occur, namely a single, clearly reportable OBE? Certainly not. If there is any uncontroversial causal relation here, it is that between the electrical stimulus and the altered brain state (the altered electrical brain processes in the angular gyrus). Between the altered brain state and the induced mental state, there is again just the well-known “correlation” that may, at best, interpret the altered state of the angular gyrus as an NCC of the OBE. But in addition to that, no further conclusions concerning brain-mind causation can be drawn from these experiments. Lesion studies form the major basis for cognitive neuropsychology, namely, the study of those alterations in human cognitive functions that result from brain lesions (McCarthy & Warrington, 1990). Here, too, we have a direct, albeit chronic, alteration of brain states (due to infarction, tumor, intracranial hemorrhage, to name just a few common causes), which results in an impairment of cognitive functions: impairment in speech understanding due to lesions in posterior language areas, disturbance of conscious visual processing due to lesions in visual cortices, impairment in object recognition (agnosia) due to lesions in posterior areas of the dominant hemisphere, and so forth. Again, the uncontroversial causal relations hold between the damaging event (e.g., a local cortical infarction) and the altered brain state. There are no further conclusions concerning brain-mind causation that could be drawn from any study from this field. For the well-known principal reasons, lesion studies cannot yield empirical evidence for an effect of the brain on the mind, they can at best demonstrate the co-occurrence of certain mental states (or functional patterns) and brain states.

Brain-Mind Causation: The Scope of Neuroscientific Evidence

Therefore, it seems that neuroscientific evidence can take us no further than to a determination of correlation or concurrence of (difficult to identify) cerebral and mental events. The very experiment presented above, however, has been used for a plea for

(quantum-level) interactionism: “How can we understand and explain the psychophysical correlations exhibited in this experiment? According to the quantum model, the conscious feelings cause the changes in brain activity to occur” (Stapp, 2007, 308). Stapp prefers interactionism in terms of quantum physics to other versions of that doctrine for two reasons. (1) Quantum interactive dualism permits the causal incompleteness of the physical, or even more: It needs “mental realities” to “complete the causal structure” (Stapp, 2007, p. 306) of the physical, and (2) it allows to specify the problematic “form of the interaction between the mentally and physically described aspects of nature” (Stapp, 2007, p. 305) in terms of a particular account of the measurement process.

Even if we take all that for granted, it is difficult to see in which sense the Ochsner et al. (2002) experiment yields specific neuroscientific evidence in support of quantum interactive dualism. In the last paragraph, I stated that the Ochsner et al. (2002) experiment is compatible with virtually any account of the mind-brain relation. Stapp (2007, p. 309) clearly anticipates this objection: “Of course, one can simply abandon the idea that ideas can actually *cause* anything physical, and view the feeling of effort as not a *cause*, but rather an *effect*, of a prefrontal excitation that causes the suppression of the limbic response, and that is caused entirely by other purely physical activities.” Here, Stapp seems to construe epiphenomenalism as a counterposition to quantum interactive dualism, and he discards that position for the often-invoked reason that epiphenomenalism “leads to the old problem: Why is consciousness present at all, and why does it feel so causally efficacious, if it has no causal efficacy at all?” (Stapp, 2007, p. 309). Stapp evokes some traditional objections to epiphenomenalism, which, under a closer look, are much less convincing than they seem to be.²⁰ But even if epiphenomenalism could be shown to be false, the Ochsner et al. (2002) experiment would leave open the possibility that the “feeling of effort” is neither a cause nor an effect, but, for example, a mental token that supervenes on the physical, or is identical with it, or just runs parallel to it, or whatever (see above). Stapp obviously has to resort to a philosophical intuition in order to find a reason why we should favor quantum interactive dualism as the theory of choice for the seeming causal interaction of the mental and the physical. Thus, we are back where we started: Neuroscience cannot, by itself, help us to make a choice between rivaling accounts of brain-mind causation. This is not to deny that the consideration of quantum processes may be relevant for an understanding of brain function and even for an understanding of the neural correlates of mental events; at least, the possibility of its relevance cannot be ruled out in advance. Like any other description of a neural correlate, however, a description of quantum-level processes can give no clue to an understanding of the relation between this cerebral event and the respective mental events.

In the preceding section, I mentioned that in addition to there being no neuroscientific data concerning the relation between mental and cerebral events, the temporal relations between these two types of events seem to be indeterminable for methodical reasons.²¹ This fact raises some problems with respect to causation in mind-brain relationships, at least if we tentatively stick to the traditional idea that effects always follow their causes, rather than preceding them or being simultaneous with them.²² First, the concept of a neural correlate of a

²⁰ See Walter (2008) for a recent rejection of some traditional objections to epiphenomenalism.

²¹ To make things even more complicated, subjective timing, the conscious experience of the temporal order of mental events, may be deceptive because the conscious experience of temporal order is but a mental “surface phenomenon” floating on a complex, underlying, unconscious, mental (and cerebral) processing of temporal relations. See Dennett & Kinsbourne (1995) for an overview of these issues.

²² Huemer and Kovitz (2003) have proposed a simultaneous theory of causation. They place causation into the context of continuously changing configurations of physical objects, where temporally extended causes are related to simultaneous and likewise temporally extended actions as their effects. They argue that this view of causation naturally arises as a result of the assumption of a continuous structure of time (and the processes that

given mental event seems to be defined in terms of simultaneousness: The neural correlate is conceived as the concurrent “counterpart” of the mental event, not as a temporally separate event that precedes the mental token (Fell et al., 2004). This would mean that if a mental token is caused by a cerebral one, then that cause can not be the neural correlate of the mental effect. On the other hand, there is possibly no way to experimentally exclude that in fact there is a minimal offset between a cerebral event and whatever we take to be its mental correlate. Hence, for the sake of a naturalistic conception of brain-mind causation, we can simply assume that there is such an unrecognizable offset, so that a certain cerebral event would minimally precede, and cause, its seeming mental “correlate” (Fell et al., 2004). Of course, the uncertainty concerning the temporal relations is symmetrical: It might just as well be that the mental event minimally precedes the cerebral one, thus causing it. This case cannot be ruled out by neuroscientific evidence, but only by an antecedent philosophical decision not to accept this reverse path (Fell et al., 2004; Huemer and Kovitz, 2003). To make things worse, in our reference experiment by Ochsner et al. (2002) it seems that “first” there is the cognitive event (the reappraisal), and “then” there is a certain activation pattern as displayed in the fMRI images. Did this seeming sequence not motivate Stapp (2007) to claim that the Ochsner et al. (2002) experiment yields evidence for mind-brain causation? Obviously, in this sort of experiment we should not confuse the independent versus dependent variable distinction with the distinction between cause and effect (or between a preceding and a following event). The seeming sequence or order of events or processes is just an artifact of the methodical frame of the respective experiment. In the Ochsner et al. (2002) experiment, the cognitive process (reappraise or attend) is the independent variable, while the cerebral process (fMRI activation) is the dependent variable. This alone, however, does neither imply a temporal sequence (the two processes must be construed as largely simultaneous) nor a direction of causation.

The Reasonable Ways of Brain-Mind Causation

Suppose, in terms of a thought experiment, that the methodical problems of the measurement of mental and cerebral processes have somehow been overcome: We are able to directly measure mental events, we have devices for recording the complete pattern of a “cerebral event”, and we can quantify and adjust the respective time windows. Still, neuroscience would provide us with nothing more than mere correlations or statements of concurrence of cerebral and mental events. Suppose further that some of these statements are law-like:²³ For any instance of a cerebral event of type C (or a token belonging to a class of well-determined cerebral events, containing C and some other events similar to C), it holds that if it occurs, it is concurrent with (or, with minimal offset, followed by) an instance of a mental event of the type M. Can we count that as “neuroscientific evidence for brain-mind causation”? Whether we will give an affirmative or a negative answer to that question depends on a philosophical (or at least non-neuroscientific) decision concerning the nature of causation. Even, however, if we decide to construe causation as nomological entailment in a very broad sense, there will be neuroscientific evidence for causation only in the very weak

occupy time) according to which even the shortest event “is already a temporally extended whole, which can be divided into indefinitely many parts...” (Huemer and Kovitz, 2003, p. 560). They hold that given that there can be no action at a temporal distance, a cause cannot precede its effect in a world with a continuous temporal structure of the type they envisage: “A continuous temporal structure requires a simultaneous view of causation” (Huemer and Kovitz, 2003, p. 561).

²³ According to Davidson’s (1980, p. 217) definition that “lawlike statements are general statements that support counterfactual and subjunctive claims, and are supported by their instances.”

sense of “*given* that causation is best construed as some sort of nomological entailment, neuroscience can yield evidence for the presence of the strict and regular brain-mind concurrence that is needed for the application of that concept of causation.” Neuroscience can only demonstrate *concurrence*, but if lawlike concurrence is all we need for stating causation, then neuroscience can give evidence for causation.²⁴

But what if we decide to choose a process (or “physical connection”) account of causation? We would then have to face two problems. (1) There seem to be no concepts that would allow us to grasp the causal interaction (the actual mechanism of causation) between a physical and a non-physical token. (2) Even if we could somehow conceptualize brain-mind causal interaction, there can be no empirical data from neuroscience that would demonstrate the presence of that interaction. This is because even ideal neuroscientific data (see above) involve only mental and cerebral tokens (and their temporal relationship), not interactions between the two. Are these problems specific for brain-mind causation, or do they hold for any causal relation in terms of process, hence also for brain-brain causation? Since most process theories are conceptualized in the terminology of the physical, they simply provide no specification for how such theories should be applied to an instance of causation that involves a non-physical relatum.²⁵ That is the sense in which problem (1) is specific for brain-mind causation,²⁶ although there is another, and perhaps more important, sense in which it is not. This is basically what Mills (1997, p. 171) called “Hume’s negative lesson,” namely, the insight that “if apprehension of some necessary connection is needed for understanding causality, mental causation poses no *special* problem, for it is at least no worse off than purely physical causation in this regard.” But this is not just an epistemological issue or an internal question of the philosophy of science. There is more to this problem than just the insight that all we can perceive or observe is sequences of events, rather than some causal power actually “at work.” After all, we are free to postulate unobservables, provided they prove to have some indispensable explanatory value in scientific theorizing (Chakravartty, 2005, p. 24). Rather, Hume’s negative lesson points to a principal problem: Any further analysis of processes or events as elements of a presumed causal relation leaves us with more detailed sequences of events, but it gives no clue as to whether there is anything that “holds these sequences together,” any causal power by virtue of which the sequence in question becomes a necessary rather than contingent one. Take the example of brain-brain causation: You can analyze the “causal chain” down to the neuronal, subneuronal, or even the quantum level, but all you will get is another micro-sequence of sub-neuronal physical events, not any information about what, if anything, connects these micro-events in the first place. There are two possible meanings of “mechanism” in the statement of problem (1) above: “Mechanism” could mean the perfectly detailed micro-sequence of events (Funkhauser, 2002, p. 337), or it could refer to

²⁴ As mentioned above, further philosophical decisions are required to preclude mind-brain causation which can, in principle, be inferred from the same neuroscientific evidence, and to exclude non-causal approaches to the relation of brain and mind, namely parallelism.

²⁵ Different process theories, though, invoke the physical realm to different degrees. Some process theories derive their terminology quite directly from physics (e.g., causation as transfer of energy or momentum, or conservation of a quantity). Other approaches apply a more neutral terminology: causation as property acquisition (Rieber 2002), causation as trope persistence (Ehring 1997, see Dowe 2007 for review). To utilize these approaches for a theory of brain-mind causation, however, one would have to refer to properties or tropes that are common to both mental and cerebral events. But at least for a non-reductionist, it is hard to conceive of properties that mental and cerebral events share. Functional properties primarily cross my mind. But if mental and cerebral events share functional properties like role-occupancy, one would perhaps say that the mental is realized by the cerebral, rather than being caused by it (see footnote 9 above).

²⁶ As Lowe (2000: 583) correctly remarks, the very term “mechanism” (let alone the term “physical connection”) is “suggestive of purely physical causation,” thus making it difficult to conceive of a “mechanism” of brain-mind causation within a non-reductive theory of the mental.

the special way in which a presumed causal power (or faculty, or capacity) “operates” within this sequence (Chakravartty, 2005, p.14). A causal reductionist would probably argue that the complete specification of the micro-sequence tells us all there is about causation, while the non-reductionist would insist that causation is the additional power that turns pure regularity into a necessary relation (see above, footnote 15, and Beebe, 2006, p. 514). Finally, the reductionist can dispense with the whole discourse of causation (Schaffer, 2008),²⁷ while the non-reductionist, no matter whether brain-brain causation or brain-mind causation is at stake, seems to be condemned to invoke a power for which there can be no direct empirical evidence.

As for problem (2) stated above, if “interaction” is understood as a causal interaction in the strong sense of a causal “mechanism,” then this problem will remain unsolved for brain-mind causation, but also for any other sort of causation (Chakravartty, 2005).

Hence, if causation is construed in terms of process or physical connection (or “transfer,” or “persistence”; see Rieber, 2002) theories, there can not even be neuroscientific evidence for causation in the very weak sense of “given that causation is best construed as some sort of process or physical connection, neuroscience can yield evidence for the presence of such a process or physical connection that is needed for the application of that concept of causation”. Again, neuroscience can only demonstrate concurrence, but if in addition to law-like concurrence, we need evidence for a process or connection for stating causation, then neuroscience cannot provide evidence for causation.²⁸

To conclude: The choice of a causal versus a non-causal interpretation of brain-mind correlations cannot be grounded in neuroscientific evidence. Neuroscience can only demonstrate concurrence of cerebral and mental events, not an additional causal relationship between them. The decision for a certain causal interpretation is based on precedent commitments to particular philosophical attitudes concerning, among other things, the mind-brain problem, scientific explanation, and the issue of reductionism. If causation is not construed in terms of a process or “physical connection” account, it may make sense to assume that cerebral events cause mental events, with all the constrictions outlined above. Current process theories are not applicable to the case of brain-mind causation, because they are largely conceptualized in terms of the physical, so that there is no conception of how they could be applied to an instance of causation that involves a non-physical relatum.

References

Beebe, H. (2006). Does anything hold the universe together? *Synthese*, 149, 509-533.

Birnbacher, D. (2006). Causal interpretations of correlations between neural and conscious events. *Journal of Consciousness Studies*, 13, 115-128.

²⁷ As Beebe (2006, pp. 514 and 527) argues, at least the regularity theorist (who holds that there is no causation in the world in addition to the contingent regularity in the sequences of events) will have to tolerate the intuitively implausible consequence that the “continued orderliness of nature” (Beebe, 2006, p. 527) is, in a sense, just luck.

²⁸ Birnbacher (2006) has argued that neuroscientific evidence is nevertheless relevant to the issue of causal interpretations of brain-mind correlations, although only in the weak sense that it helps to confirm or disconfirm certain interpretations if basic “methodological” principles like parsimony, simplicity, and coherence are considered in addition to the mere empirical evidence provided by neuroscience. Birnbacher favors a causal interpretation of brain-mind relations in terms of epiphenomenalism.

- Blanke, O., Ortigue, S., Landis, T., & Seeck, M. (2002). Stimulating illusory own-body perceptions. *Nature*, *419*, 269-270.
- Brown, G.G., Eyler, L.T. (2006). Methodological and conceptual issues in functional magnetic resonance imaging: applications to schizophrenia research. *Annual Review of Clinical Psychology*, *2*, 51-81.
- Chalmers, D. (2000). What is neural correlate of consciousness? In T. Metzinger (Ed.), *Neural Correlates of Consciousness: Empirical and Conceptual Questions*. Cambridge, MA: MIT Press.
- Collins, J., Hall, N., & Paul, L.A. (2004). Counterfactuals and causation: history, problems, and prospects. In J.D. Collins, E.J. Hall, L.A. Paul (Eds.), *Causation and counterfactuals* (pp. 1-57). London: Bradford Books.
- Chakravartty, A. (2005). Causal realism: events and processes. *Erkenntnis*, *63*, 7-31.
- Davidson, D. (1980). Mental events. In D. Davidson, *Essays on actions and events* (pp. 207-225). Oxford: Clarendon Press.
- Dennett, D.C., Kinsbourne, M. (1995). Time and the observer: the where and when of consciousness in the brain. *Behavioral and Brain Sciences*, *15*, 183-247.
- Dowe, P. (2007). Causal processes. In *Stanford Encyclopedia of Philosophy*. Retrieved January 1st, 2009, from <http://plato.stanford.edu/entries>.
- Dretske, F. (1988). *Explaining behavior. Reasons in a world of causes*. Cambridge, MA: MIT Press.
- Ehring, D. (1997). *Causation and persistence*. Oxford: Oxford University Press.
- Esfeld, M. (2007). Kausalität. In A. Bartels & M. Stöckler (Eds.), *Wissenschaftstheorie. Ein Studienbuch* (pp. 89-107). Paderborn, Germany : Mentis.
- Fell, J., Elger, C.E., & Kurthen, M. (2004). Do neural correlates of consciousness cause conscious states? *Medical Hypotheses*, *63*, 367-369.
- Freeman, W.J. (1999). Consciousness, intentionality, and causality. *Journal of Consciousness Studies*, *6*, 143-172.
- Funkhouser, E. (2002). Three varieties of causal overdetermination. *Pacific Philosophical Quarterly*, *83*, 335-351.
- Henson, R. (2005). What can functional neuroimaging tell the experimental psychologist? *Quarterly Journal of Experimental Psychology A*, *58*, 193-233.
- Huemer, M., & Kovitz, B. (2003). Causation as simultaneous and continuous. *The Philosophical Quarterly*, *53*, 556-565.
- Jaworski, W. (2006). Mental causation from the top down. *Erkenntnis*, *65*, 277-299.
- Kim, J. (1996). *Philosophy of mind*. Boulder, CO: Westview.
- Knight, R.T. (2007). Neural networks debunk phrenology. *Science*, *316*, 1578-1579.
- Lowe, E.J. (2000). Causal closure principles and emergentism. *Philosophy*, *75*, 571-585.
- Lowe, E.J. (2006). Non-cartesian substance dualism and the problem of mental causation. *Erkenntnis*, *65*, 5-23.
- MacDonald, G. (2007). Emergence and causal powers. *Erkenntnis*, *67*, 239-253.
- Marcus, E. (2005). Mental causation in a physical world. *Philosophical Studies*, *122*, 27-50.
- McCarthy, R.A., & Warrington, E.K. (1990). *Cognitive Neuropsychology*. San Diego, CA: Academic Press.

- McLaughlin, B. (2007). Mental causation and Shoemaker-realization. *Erkenntnis*, 67, 149-172.
- McLaughlin, B., & Bennett, K. (2005). Supervenience. In *Stanford Encyclopedia of Philosophy*. Retrieved January 1st, 2009, from <http://plato.stanford.edu/entries>.
- Metzinger, T. (2000). *Neural correlates of consciousness: Empirical and conceptual questions*. Cambridge, MA: MIT Press.
- Mills, E. (1997). Interactionism and physicality. *Ratio*, 10, 169-183.
- Noe, A & Thompson, E. (2004). Are there neural correlates of consciousness? *Journal of Consciousness Studies*, 11, 3-28.
- Ochsner, K.N., Bunge, S.A., Gross, J.J., & Gabrieli, J.D.E. (2002). Rethinking feelings: an fMRI study of the cognitive regulation of emotion. *Journal of Cognitive Neuroscience*, 14, 1215-1229.
- Papineau, D. (1993). *Philosophical naturalism*. Oxford: Blackwell.
- Polger, T.W. (2007). Realization and the metaphysics of mind. *Australasian Journal of Philosophy*, 85, 233-259.
- Putnam, H. (1975). *Mind, language, and reality* (Vol. 2, chap 21, pp. 429-440). Cambridge, UK: Cambridge University Press.
- Rieber, S.D. (2002). Causation as property acquisition. *Philosophical Studies*, 109, 53-74.
- Schaffer, J. (2001). Causes as probability-raisers of processes. *Journal of Philosophy*, 98, 75-92.
- Schaffer, J. (2007). The metaphysics of causation. In *Stanford Encyclopedia of Philosophy*. Retrieved January 1st, 2009, from <http://plato.stanford.edu/entries>.
- Schaffer, J. (2008). Causation and laws of nature: reductionism. In T. Sider, J. Hawthorne, & D. W. Zimmerman (Eds.), *Contemporary debates in metaphysics* (pp. 82-108). Malden, MA: Blackwell.
- Stapp, H.P. (2005). Quantum interactive dualism. An alternative to materialism. *Journal of Consciousness Studies*, 12, 43-58.
- Stapp, H.P. (2007). Quantum mechanical theories of consciousness. In M. Velmans, & S. Schneider (Eds.), *The Blackwell companion to consciousness* (pp. 300-312). Malden, MA: Blackwell.
- Uttal, W.R. (2003). *The new phrenology: The limits of localizing cognitive processes in the brain*. Cambridge, MA: MIT Press.
- Walter, S. (2008). Ist der Epiphänomenalismus absurd? Ein frischer Blick auf eine totgeglaubte Position. *Zeitschrift für philosophische Forschung*, 62, 415-432.

Figure 1:

**Cognitive neuroscience:
General architecture of the experiment**

