Interventionist Mental Causation and the Methods of Cognitive Neuroscience

Mario Günther

January 3, 2018

This is a Draft. Please Do Not Circulate, Quote or Cite

Contents

1 Introduction 2

2 Causation and Methods in Cognitive Neuroscience 4

3 Woodward’s Interventionist Account of Causation 5

4 Method I, Baumgartner’s Causal Exclusion Argument and Woodward’s Reply 9

5 Method II and Causal Exclusion 11

6 Discussion 13

7 Conclusion 15

Abstract 1
We impose properties of causation, as assumed in cognitive neuroscience, upon Woodward (2005, 2015)’s account of interventionism. Within the resulting framework, we investigate to what extent we are justified to derive causal relations between mental properties and properties of the brain, if certain methods are used in the neuroscientific studies.

**Keywords.** Mental Causation, Interventionism, Methods of Cognitive Neuroscience, Causal Exclusion Arguments, Supervenience.

1 **Introduction**

Should cognitive neuroscientists interpret the results of their studies as establishing genuine causal relations? Well, the answer depends (at least) on three factors.

1. What does ‘causal relation’ mean?
2. Which methods are employed in the studies?
3. How is the relation between the mental and the brain?

Woodward (2005)’s interventionist notion of causation captures, or so we argue, the notion of causation employed in cognitive neuroscience quite well. Furthermore, we coarsely distinguish between two methods: one manipulates a mental property and measures a property of the brain, the other proceeds vice versa.

The search for the neural correlates of mental functions is premised upon the idea that there is some dependency between mental properties and properties of the brain.¹ For reasons of cautiousness, we assume a minimal dependency relation in this paper, which we call minimal supervenience.² We represent properties by variables, as in the following definition.

**Definition 1. Minimal Supervenience**

A variable $M$ minimally supervenes on variable $P$ (for $M \neq P$) iff

---

¹Cf. Squire et al. (2008, Ch. 53).
²In a follow-up paper, we investigate what happens to the results obtained here if different relations between the mental and the brain are assumed, such as identity, functional reduction, causal relations, etc.
(i) $M$ and $P$ occur synchronically, and

(ii) any change of the value of $M$ necessarily changes the value of $P$.

Minimal supervenience is defined by the conditions that are common to most supervenience relations. Hence, it expresses necessary conditions for supervenience, which are – depending on the philosophical view – not necessarily sufficient. Nevertheless we use ‘minimally supervenes’ and ‘supervenes’ interchangeably, unless noted otherwise.

Finally, we assume that brain properties are causes of other brain properties. We obtain the picture of Figure 1, which resembles Kim (2005)’s canonical diagram. Given the depicted assumptions, we argue that cognitive neuroscientists should interpret the findings of their studies as establishing genuine causal relations.

![Figure 1: A causal network extended by minimal supervenience relations. The undirected edges stand for these supervenience relations. Note that there is a causal relation between brain property $P_1$ and brain property $P_2$.](image)

In Section 2, we introduce the notion of causation assumed in cognitive neuroscience and the two methods we distinguish. In Section 3, Woodward (2005)’s interventionist account of causation for models of exclusively causal relations is presented. In Section 4, we review Baumgartner (2009)’s causal exclusion argument for the first method and Woodward (2015)’s reply. In Section 5, we find that Baumgartner’s argument does not apply with respect to the second method on Woodward’s original account. We discuss the results in Section 6.
2 Causation and Methods in Cognitive Neuroscience

The research in cognitive neuroscience assumes a certain notion of causation. This notion is characterised by a temporal order: a cause precedes its effect in time. By Definition 1, condition (i), if a variable $M$ supervenes on $P$, then neither $M$ is a cause of $P$, nor $P$ is a cause of $M$. For neither temporally precedes the other. Furthermore, causal relations are assumed to generate statistical dependencies. According to this theoretical assumption, if a cognitive neuroscientist manipulates the cause of an effect, she enforces a statistical dependency between them. Textbooks on experimental research widely agree that two conditions must be met to show a causal relation between the variable $X$ and the variable $Y$:

$(c_i)$ A systematic variation of $X$‘s value changes $Y$‘s value, and simultaneously

$(c_{ii})$ all other variables are controlled for to exclude other possible causes of $Y$.\(^3\)

On the face of it, the notion of causation assumed in cognitive neuroscience, including the corresponding experimental practice, seems to be well-captured by James Woodward (2005)’s interventionist account of causation. We will have a closer look at this claim in the next section.

Many of the currently conducted studies in cognitive neuroscience adhere to one of two methods.

Method I: Manipulate a mental property of a participant and measure, subsequently, one of her brain properties.

Method II: Manipulate a brain property of a participant and inquire, subsequently, one of her mental properties.

In studies employing Method I, the mental property of a participant is manipulated by presenting different task conditions. Afterwards, the changes in the participant’s neural activity are measured. Examples are studies in which changes in

\(^3\)See, for example, Carter and Shieh (2015); Gravetter and Forzano (2011); Squire et al. (2008); Windhorst and Johansson (1999)
neural activity are measured by means of Functional Magnetic Resonance Imaging (fMRI) or Electro-Encephalogram (EEG). Method I studies are often interpreted to merely show a correlation rather than a genuine causal relation.

In studies employing Method II, a brain property of a participant is manipulated by brain stimulation techniques. Thereafter, the changes in the participant’s mental property are inquired (or ‘measured’). Examples are studies in which the electrical activity of the brain is directly manipulated by brain stimulation techniques such as Deep Brain Stimulation (DBS) and Transcranial Magnetic Stimulation (TMS). DBS has been used, e.g., to treat a variety of intractable pain syndromes, including neuropathic pain, phantom-limb pain, failed low back pain, and cluster-headache pain.\(^4\) We take the pain relief to be a mental variable, which we can simply ‘measure’ by a participant’s report. TMS is a noninvasive procedure that uses magnetic fields to stimulate nerve cells in the brain. Like DBS, TMS can be used to improve symptoms of depression.\(^5\) We take the mood reported by a participant to be an inquirable mental variable. In contrast to Method I studies, Method II studies are normally interpreted as establishing causal relations.\(^6\)

Whether we are allowed to derive causal relations from studies in cognitive neuroscience depends on the question what causal relations are. We turn now to a recent answer due to Woodward (2005).

### 3 Woodward’s Interventionist Account of Causation

Woodward (2005) provides the following definitions of direct and contributing cause, which we slightly rephrased.

**Definition 2. Direct Cause (cf. Woodward (2005))**
The variable \(X\) is a direct cause of the variable \(Y\) relative to a variable set \(\mathcal{V}\) iff there is a possible intervention on \(X\) that changes \(Y\)’s value, presupposed the values of all other variables \(V_i\) in \(\mathcal{V}\) are kept fixed.\(^7\)

---


\(^5\)See George et al. (2003).


\(^7\)For simplicity, we omitted “or \(Y\)’s probability distribution” after “\(Y\)’s value”. We continue to do so as no argument in the paper hinges on it.
We observe that an intervention on \( X \) that changes the value of \( Y \) corresponds to \((c_i)\), and the presupposition that the values of all other variables \( V_i \) in \( V \) are kept fixed formally implements \((c_{ii})\). Hence, Woodward’s definition of direct cause formally captures the necessary conditions of cognitive neuroscience’s notion of causation.

**Definition 3. Contributing Cause (cf. Woodward (2005))**
The variable \( X \) is a contributing cause of \( Y \) relative to a variable set \( V \) iff

(i) there is a directed path from \( X \) to \( Y \) such that each link in this path is a direct causal relation and

(ii) there is some intervention on \( X \) that changes \( Y \), presupposed the values of all variables in \( V \) that are not on the path are kept fixed.

From now on, we say that \( X \) is a cause of \( Y \) iff \( X \) is a contributing cause. Note that direct causes are a limiting case of contributing causes for a path of length 1.

Both of the definitions cite the term ‘intervention’. In the context of experimental design, an intervention is a manipulation that changes the value of the (independent) variable \( X \). Woodward defines interventions by means of an intervention variable.

**Definition 4. Intervention Variable (cf. Woodward (2005))**
Let \( I, X, Y \) be variables and \( \{I, X, Y\} \subseteq V \) be a set of variables. \( I \) is an intervention variable for \( X \) with respect to \( Y \) relative to \( V \) iff

(1) Setting \( I = i \) determines \( X = x \).

(2) \( I \) breaks \( X \)’s causal dependences on all variables in \( V \setminus I \).

(3) Any directed path in \( V \) from \( I \) to \( Y \) includes \( X \).

(4) \( I \) is statistically independent of any variable \( V \in V \setminus I \) that is

\[(a) \quad \text{a cause of } Y, \quad \text{and}\]

\[\text{Corresponding to the previous footnote, we omit “or } P(X = x), \text{ where } P \text{ is a probability measure” here.}\]
An intervention is thus, according to Woodward’s formalism, a setting of the intervention variable to a value \( I = i \) that changes the value of the variable \( X \). By the conditions (1) and (2), \( I \) is the only and a direct cause of \( X \). Furthermore, Woodward demands that such an intervention must satisfy the conditions (1)–(4) to establish a causal relation. We visualize an intervention variable and its properties in Figure 2 and Figure 3, respectively.

![Figure 2: A directed acyclic graph \( G = \langle V, A \rangle \) representing a causal network, whose sets of variables and arrows are \( V = \{I, X, Y\} \) and \( A = \{\langle I, X \rangle, \langle X, Y \rangle\} \), respectively. Each directed arrow represents a causal relation, or equivalently causal dependency. A directed arrow with question mark indicates which potential causal relation is investigated. Here the candidate under investigation is the relation between \( X \) and \( Y \). By Definition 4, \( I \) is an intervention variable for \( X \) with respect to \( Y \) relative to \( V \).]
Figure 3: (a), (b), and (c) illustrate the conditions (2), (3), and (4) of Definition 4, respectively. A crossed arrow represents that a causal dependency cannot exist, if the respective conditions are satisfied. A dashed line stands for statistical independence. Note that no question mark arrows, crossed arrows and/or dashed lines are elements of basic causal networks. Those types of edge serve illustrative purposes only.

Definition 4 is meant to exclude potential confounding variables, which need to be controlled for in an experiment. This intention behind an intervention variable will be crucial to the debate on which variables ought to be controlled for in assessing causal relations, especially with respect to the causal efficacy of mental properties. If Definition 4 is modified with respect to the set of variables that should be kept fixed, we obtain a different (but still interventionist) notion of cause.

Woodward’s framework is intended for models of exclusively causal relations, i.e. the variables in such a model may be causally related or correlated, but they do not stand in a supervenience relation. In such a model, each represented variable is independently fixable, i.e. it is possible to set the value of each variable independently of the other represented variables. In a model including a supervenience relation, independent fixability is violated. For an intervention on the supervening variable necessarily changes its subvenient variable. This prompts the question how to read the Definitions 2, 3, and 4 for models including causal and supervenience relations.
4 Method I, Baumgartner’s Causal Exclusion Argument and Woodward’s Reply

We model now Method I under Baumgartner (2009)’s reading of Woodward (2005)’s interventionist account of causation extended by minimal supervenience relations and show that Method I falls prey to Baumgartner (2009)’s Causal Exclusion Argument.

We extend Woodward’s interventionist account by including supervenience relations, as defined by Definition 1. The result are models containing two types of relations: supervenience relations between properties of an event (e.g. between $M_1$ and $P_1$) and causal relations between events (e.g. the event $M_1/P_1$ is a cause of $M_2/P_2$.) For such models, Baumgartner does not modify the Definitions 2, 3, and 4.

Method I studies examine the effect on brain state $P_2$ after manipulating the mental state $M_1$ (see Figure 4). Typically, a cognitive neuroscientist, let’s call her Pat, manipulates the mental state $M_1$ of the participant in an experiment: the subject faces different task or stimulus conditions which are supposed to induce changes in $M_1$. Then, Pat investigates the effect of her manipulations by measuring the participant’s brain state $P_2$ in all conditions. Using controls like randomization of subjects, our ideal neuroscientist Pat ensures that the conditions only differ on the manipulated mental variable. If Pat finds a significant difference in brain activity between the conditions, she derives that $M_1$ has had an effect on $P_2$.

![Figure 4](image_url)

Figure 4: The relation investigated by studies of cognitive neuroscience using Method I.
However, Method I studies are normally interpreted to show a correlation only. Are we justified in deriving a causal relation from Method I studies within the framework of Woodward’s extended interventionism? Baumgartner (2009) says no. More specifically, Baumgartner (2009) provides a causal exclusion argument that says: we are not allowed to derive a causal relation between $M_1$ and $P_2$, if we assume Woodward’s interventionism extended by supervenience relations.

Baumgartner (2009)’s causal exclusion argument runs as follows. Woodward’s interventionism entails the implication if $M_1$ is a cause of $P_2$ relative to $\mathcal{V} = \{M_1, M_2, P_1, P_2\}$, then there is an intervention variable $I$ possible that changes the value of $M_1$ and is statistically independent of any variable $V \in \mathcal{V} \setminus I$ that is (a) a cause of $P_2$ and (b) on a directed path not including $M_1$ (compare p. 170). But such an $I$ is impossible. Since $M_1$ minimally supervenes on $P_1$, any $I$ that changes the value of $M_1$ also changes the value of $P_1$. Hence, $I$ is statistically dependent on $P_1$. Moreover, $P_1$ is (a) a direct cause of $P_2$ by assumption and (b) on a directed path not including $M_1$. If $I$ is an intervention variable, then – by condition (4) of Definition 4 – $I$ is required to be statistically independent of $P_1$. Therefore, $I$ is no intervention variable. By modus tollens on Baumgartner’s implication, $M_1$ is not a cause of $P_2$.

In natural language, if Pat intervenes on a participant’s mental state, she simultaneously intervenes on the physical state due to the supervenience relation between $M_1$ and $P_1$. Because of this non-causal relation, Pat cannot control the effect of $P_1$ on $P_2$ when manipulating $M_1$. By condition (i) of Definition 3, $M_1$ is also no contributing cause of $P_2$, since the link between $M_1$ and $P_1$ is no causal relation. Poor Pat, $P_1$ is a confounding variable for which she cannot control. Baumgartner’s reading of Woodward’s extended interventionism recommends that we should not derive a causal relation between $M_1$ and $P_2$ – quite in agreement with the current interpretation of Method I studies.

In reply to Baumgartner’s argument, Woodward (2015) provides an interventionist framework for causal relations in the presence of supervenience relations. He emphasizes that Woodward (2005)’s interventionist account of causation, as given by Definitions 2, 3, and 4, is intended for models of exclusively causal relations. If we want to include non-causal relations, such as the minimal supervenience of Definition 1, we need to modify the interventionist account. In light of Baumgartner’s argument, Woodward (2015)’s modification idea is to relax the requirement, or presupposition, to keep the variables fixed. In models including non-causal
relations, only ‘appropriate’ variables need to be fixed for establishing a causal relation. When assessing a causal relation between a supervening variable and another variable, for example, it be not appropriate to keep fixed its supervenience base(s). In the modified framework, he claims, supervening properties can be causally efficacious.

To be more precise, Woodward (2015) modifies the notion of an intervention variable. Such an intervention variable $I^*$ on $X$ with respect to $Y$ (relative to $V$) (a) respects the supervenience relation between $X$ and its supervenience base $SB(X)$, if there is one, and (b) requires that the conditions (1)-(4) of Definition 4 are restricted to causally related or correlated variables, but do not apply to supervenient and subvening variables. By (a), an $I^*$-intervention on $X$ is thus at the same time an $I^*$-intervention on $SB(X)$, if such a supervenience base exists.

Consider Figure 4 again. Applying the modified Definition 2, we obtain that $M_1$ is a direct cause of $P_2$ relative to $V = \{M_1, M_2, P_1, P_2\}$ iff there is a possible $I^*$-intervention on $M_1$ that changes the value of $P_2$, presupposed the values of all other appropriate variables in $V$ are kept fixed. Since $P_1$ is the supervenience base of $M_1$, $P_1$ is not one of the appropriate variables to keep fixed. Hence, $M_1$ is a cause of $P_2$ according to Woodward’s modified account.

As we have seen, Baumgartner (2009)’s causal exclusion argument entails that his reading of Woodward (2005)’s original interventionism does not recommend to interpret the results of Method I studies as establishing causal relations. In contrast, Woodward (2015)’s modified interventionism allows us to derive causal relations between mental states and brain states from Method I studies – somewhat at odds with the current interpretation in cognitive neuroscience.

## 5 Method II and Causal Exclusion

We revisit now Baumgartner (2009)’s Causal Exclusion Argument for his and Woodward (2015)’s reading of interventionism (both extended by minimal supervenience relations) with respect to Method II studies. Those studies examine the effect on mental state $M_2$ after manipulating the brain state $P_1$ (see Figure 5). Our ideal cognitive neuroscientist Pat manipulates the physical state $P_1$ of the participant’s brain by directly intervening upon the electrical activity in the
brain using brain stimulation techniques. Then, Pat investigates the effect of her manipulations by inquiring the participant’s mental state $M_2$. Using controls like sham stimulation, she ensures that the experimental conditions – often stimulation vs. no stimulation – only differ with respect to $P_1$.\(^9\) If Pat finds a significant difference in the participant’s mental state, she derives that $P_1$ has had an effect on $M_2$.

![Diagram](image)

Figure 5: The relation investigated by studies of cognitive neuroscience using Method II.

Method II studies are often interpreted to show a proper causal relation.\(^10\) Are we justified in deriving a causal relation from those studies according to Woodward’s original account? As it turns out, yes.

We obtain an implication similar to Baumgartner (2009)’s: if $P_1$ is a direct cause of $M_2$ relative to $\mathcal{V} = \{M_1, M_2, P_1, P_2\}$, then there is a variable $I$ possible that changes the value of $P_1$ and is statistically independent of any variable $V \in \mathcal{V}$ that is (i) a cause of $M_2$ and (ii) on a directed path not including $P_1$. This time such an $I$ is possible. The reason is that there is no variable apart from $P_1$ that can be a cause of $M_2$ and is on a directed path not including $P_1$. Since there is a supervenience relation between $M_2$ and $P_2$, the only other candidate cause for $M_2$ is $M_1$. But Baumgartner’s reading denies that $M_1$ can be a cause of $M_2$. For an intervention on $M_1$ does not allow to keep fixed its supervenience base $P_1$, as required by Baumgartner. Pictorially speaking, $P_1$ is on any directed path to $M_2$.

\(^9\) Sham stimulation is a generic term to indicate an inactive form of stimulation (e.g., a very brief or weak one) that is used in research to control for the placebo effect. The subject believes he/she is being stimulated normally, but there should not be any real effects.

\(^{10}\) See, for example, Martin and Gotts (2005) and Romei et al. (2012).
if there is one. Hence, a causal exclusion argument in Baumgartner’s style does not apply with respect to Method II such that his reading of Woodward (2005) allows us to derive causal relations from studies using brain stimulation.

Interestingly, Woodward (2015)’s reading faces a putative problem with respect to Method II studies. The potential problem is that $P_1$ gets a competitor candidate cause of $M_2$, viz. $M_1$. Recall that Woodward’s modification implies that supervenience bases need not be kept fixed. According to the modified account, we obtain the implication: if $M_1$ is a cause of $M_2$ relative to $V = \{M_1, M_2, P_1, P_2\}$, then there is a variable $I'$ possible that changes the value of $M_1$ and is statistically independent of any appropriate variable $V \in V$ that is (i) a cause of $M_2$ and (ii) on a directed path not including $M_1$. Since the supervenience bases of $M_1$ and $M_2$, i.e. $P_1$ and $P_2$ respectively, do not belong to the appropriate variables, there is such an $I'$ possible that changes the value of $M_1$ and, perhaps, ‘miraculously’ results in a change of $M_2$. Hence, $M_1$ is an alternative candidate cause of $M_2$. The putative problem on the modified account is thus that we cannot know from a Method II study whether $P_1$ or $M_1$ is the cause of $M_2$. We briefly discuss this issue in the next section.

6 Discussion

Consider again Figure 4. Woodward (2015)’s account draws an arrow from $M_1$ to $P_2$ in addition to the assumed arrow from $P_1$ to $P_2$, if the value of $P_2$ changes when intervening on $M_1$. However, the arrow from $M_1$ to $P_2$ does not mean that $M_1$ has an effect on $P_2$ over and above $P_1$’s effect on $P_2$. Similarly, if interventions on $M_1$ change $M_2$’s value, then Woodward (2015) draws an arrow from $M_1$ to $M_2$ but also from $M_1$ to $P_2$, as $P_2$’s value will change under an intervention on $M_1$ that changes $M_2$.

---

11Notice that the possible intervention on $P_1$ does not necessarily change $M_1$ due to the supervenience relation, but it does change $P_2$ by the assumed causal relation. However, $P_2$ is not a cause of $M_2$ due to their supervenience relation, and thus $P_1$ is a direct cause of $M_2$, given a change of $M_2$’s value has been detected.

12Notice the possibility, however, that an intervention on $M_1$ changes the value of $P_2$ without changing the value of $M_2$. In such a case $M_1$ disqualifies as a cause of $M_2$, while $M_1$ is still a cause of $P_2$. 

13
It seems that Baumgartner (2009)’s Causal Exclusion Argument rests on the intuition that $M_1$ can only be causally efficacious, if it has a causal influence on $M_2$ or $P_2$ which is separate from $P_1$’s causal influence on those variables. Due to this implicit premise, or so it seems, Baumgartner thinks $P_1$ should be kept fixed in order to test the ‘independent’ causal influence of $M_1$. Woodward (2015) simply does not share this intuition and Baumgartner does not provide any argument why we should adopt it.

The question which variables should be controlled for when assessing the causal efficacy of the mental is at the heart of Baumgartner (2009)’s Causal Exclusion Argument. Should a cognitive neuroscientist adopt Baumgartner’s intuition and control for $P_1$ when intervening on $M_1$? No, we shouldn’t ask scientists for the impossible. Instead, we should encourage neuroscientists to shed light on the relation between the mental and the physical. The more they know the specific relation, the clearer they can see that an intervention on $M_1$, for instance, is ‘automatically’ an intervention on its supervenience base $P_1$. In this sense, an intervention on $M_i$ should respect the supervenience relation between $M_i$ and $SB(M_i) = P_i$, i.e. setting $M_i = m_i$ specifies $P_i = p_i$ that is consistent with $M_i = m_i$ according to the supervenience relation.

To figure out the details which mental properties supervene on which physical properties is of utmost importance, as the following consideration suggests. In Figure 1, $M_1$ supervenes on $P_1$. Hence, it is possible to change the value of $P_1$ without changing the value of $M_1$ (but not vice versa). Thus we may empirically figure out the values of $P_1$ under which $M_1$ remains invariant in order to test whether $M_1$ is a cause of $M_2$ in a sense, in which $P_1$ is not. If we intervene on $P_1$ such that $M_1$ keeps its value and $M_2$ as well, but intervening on $M_1$ changes $M_2$’s value, then $M_1$ is a cause in a sense in which $P_1$ is not.\(^{13}\)

Consider again Figure 5. Assume it is empirically found that $P_1$ is a cause of $M_2$ and further empirical investigations establish that $M_1$ is a cause of $M_2$. This is a putative problem for Woodward (2015)’s account. The cause of $M_2$ is underdetermined: $M_1$ is a cause of $M_2$ and $P_1$ is a cause of $M_2$. However, this is

\(^{13}\)Note that the possibility to experimentally distinguish the causal influence of $P_1$ and $M_1$ depends on the possibility to change the supervenience base $P_1$ without changing the supervenient variable $M_1$. These considerations become especially pertinent in the sequel paper when relations of multiple realizability are considered. For an attempt in this direction, see List and Menzies (2009).
only a putative problem. According to Definition 3, both $M_1$ and $P_1$ may qualify as contributing causes. Furthermore, Woodward (2005, 2015) explicitly states that arrows do not necessarily correspond to (physical) causal mechanisms which can be distinguished. Hence, the underdetermination problem is no problem for Woodward’s formalism.

Apart from the misunderstanding of Woodward’s formalism, the underdetermination raises the question whether there are reasons to distinguish between the supervenient variable $M_i$ and its supervenience base $P_i$ with respect to causal efficacy. After all, $M_i/P_i$ is one event. We think this is a contingent matter which has no universally valid answer. Rather it illustrates once more why a positive characterization of the relation between the mental and the brain is desirable.

7 Conclusion

We imposed properties of causation, as assumed in cognitive neuroscience, upon Woodward (2005, 2015)’s account of interventionism, and extended it by minimal supervenience relations. Within the resulting framework, we investigated to what extent we are justified to derive causal relations between mental properties and properties of the brain from methods used in studies of cognitive neuroscience.

The results of Baumgartner (2009)’s reading of Woodward (2005)’s interventionist framework extended by a minimal supervenience relation fits the current interpretations of the results in cognitive neuroscience. Correspondingly, studies that manipulate mental properties and measure properties of the brain do not allow us to derive causal relations. In contrast, Baumgartner (2009)’s reading justifies to derive causal relations from studies which manipulate brain properties and inquire mental properties. Those results agree with the current interpretations in cognitive neuroscience. If Baumgartner (2009)’s reading is correct, this provides support for the interpretations. If the interpretations of cognitive neuroscientists are justified, this provides support for Baumgartner (2009)’s reading.

Woodward (2015) modifies his original framework to include non-causal supervenience relations. On this account, we are allowed to derive genuine causal relations from both methodological kinds of study. We have seen that the idea behind Baumgartner (2009)’s causal exclusion argument is if the property $M_1$ is a cause
of $P_2$ (or $M_2$), then $M_1$ must have a causal impact on $P_2$ (or $M_2$) while keeping fixed $P_1$’s causal impact on those variables. However, as Woodward (2015) notes, it seems inappropriate to control for the supervenience base $SB(M_1) = P_1$ when assessing the causal efficacy of $M_1$. We side with Woodward, as it doesn’t seem to be a necessary condition for a cause that we can intervene upon it while keeping fixed its supervenience base. The reason is simply that this is impossible. If it were a necessary condition for causation, then we would never have mental causation, which is absurd.

So far we see no reasons why we shouldn’t adopt Woodward (2015)’s extended interventionism, enriched by properties of causation as assumed in cognitive neuroscience, as the theoretical framework for determining whether the corresponding studies establish genuine causal relations. On the one hand, Baumgartner (2009) does not provide any positive argument against Woodward (2015)’s account. On the other hand, our resulting framework is formally precise and provides unique answers, is easily applicable and fits nicely to the experimental practice of cognitive neuroscientists. Therefore, we recommend – against the current interpretations – that cognitive neuroscientists should dare to interpret both methodological kinds of studies as establishing genuine causal relations.

The results and our recommendation depend on our assumption of a minimal supervenience relation between mental properties and brain properties. While most cognitive neuroscientists agree that there is at least such a minimal dependence relation between mental processes and brain processes, there could also be a stronger relation such as multiple realizability, functional reduction, or even type identity. Of course, minimal supervenience provides no satisfaction in contrast to a positive account of the relation between the mental and the brain. The investigation of how the obtained results change when we spell out the mind-brain relation is left for future work.

Acknowledgments. Thanks to Borut Trpin for helpful comments. I am grateful that I had the opportunity to present parts of this paper at the ECAP9 Symposium on “Current Trends in Neurophilosophy”. This research is supported by the Graduate School of Systemic Neurosciences.
References


