

# Mental causation, interventionism, and probabilistic supervenience<sup>1</sup>

Alexander Gebharter \* Maria Sekatskaya

**Abstract:** Mental causation is notoriously threatened by the causal exclusion argument. A prominent strategy to save mental causation from causal exclusion consists in subscribing to an interventionist account of causation. This move has, however, recently been challenged by several authors. In this paper, we do two things: We (i) develop what we consider to be the strongest version of the interventionist causal exclusion argument currently on the market and (ii) propose a new way how it can in principle be overcome. In particular, we propose to replace strict supervenience in the assumption that the mental supervenes on the physical by probabilistic supervenience and show how this move has the potential to license the inference to mental causation. Finally, we argue that probabilistic supervenience captures some of the most important intuitions that strict supervenience captures and discuss possible objections to weakening strict supervenience in the way we suggest.

*Key words:* causal exclusion, interventionism, mental causation, supervenience

---

<sup>1</sup> This is a draft paper. Please do not cite or quote without permission. The final version of this paper is published under the following bibliographical data: Gebharter, A., & Sekatskaya, M. (2024). Mental causation, interventionism, and probabilistic supervenience. *Synthese*, 203, 206. doi: 10.1007/s11229-024-04608-w

1

## 2 **1 Introduction**

3 Intuitively, it seems plausible that one's mental states are not identical to their physical  
4 realizers and have some autonomy *vis-à-vis* the latter. Mental states make a difference for  
5 how the future unfolds, which allows us to interact with our environment and gives us to  
6 some extent control over our lives. However, mental causation is threatened by several  
7 philosophical objections. One of the most prominent is the causal exclusion argument (Kim,  
8 2005). In a nutshell, the argument says that if the mental is ontologically non-identical to the  
9 physical, the mental supervenes on the physical, for every physical state there is a sufficient  
10 physical cause, and causal overdetermination does not happen systematically, then the  
11 mental is causally inert.

12 Supporters of mental causation have several options to block<sup>2</sup> the exclusion argument.  
13 One can either attack (A) one or several of the argument's premises or (B) the argument's  
14 validity. One way to contest the truth of the premises or the argument's validity consists in  
15 further specifying the notion of causation used in the argument (cf. Hitchcock, 2012). This  
16 paper builds on the debate that resulted from one such attempt, in particular, from the  
17 attempt to escape causal exclusion by subscribing to an interventionist account of causation  
18 such as Woodward's (2003). Some of the advantages of interventionism, if compared to  
19 other theories of causation, is that it is close to scientific practice, comes with a clear  
20 semantics for causation as well as a sophisticated methodology for causal inference, and  
21 provides a rich and easy to expand basis for exploring a multitude of philosophical issues

---

<sup>2</sup> By *blocking* the exclusion argument we mean to render mental causation conceptually compatible with the truth of the other premises used in the argument. Thus, blocking the exclusion argument does not yet guarantee that there actually is mental causation. It rather paves the road for that by establishing the possibility of mental causation given the premises are true. This is what the present paper is about.

1 directly related to causation.<sup>3</sup> We do not go into further details why interventionism is an  
2 especially promising candidate for reconstructing the causal exclusion argument but rather  
3 content ourselves with pointing the reader to the fact that there exists a rich literature about  
4 this specific topic and let this fact speak for itself (see, e.g., Baumgartner, 2010, 2013, 2018;  
5 Eronen & Brooks, 2014; Gebharter, 2017a; Hoffmann-Kolss, 2014, 2021; Kinney, 2023;  
6 Stern & Eva, 2023; Woodward, 2015).

7 Because much has been written about causal exclusion and interventionism, we need to  
8 narrow down the scope of this paper even further. In particular, we will only consider what  
9 in our view is the most threatening objection to interventionist mental causation currently on  
10 the market. This threat emerged from a series of objections raised by authors such as  
11 Baumgartner (2010, 2013), Baumgartner and Gebharter (2016), Gebharter (2017a, 2017b),  
12 and culminated in Baumgartner's (2018) "The inherent empirical underdetermination of  
13 mental causation". Our main focus will not lie on any particular argument to be found in these  
14 papers, but rather on what we take to be the strongest argument based upon these  
15 arguments. In a nutshell, the argument we will develop aims at establishing that a mental  
16 state causing some other mental or physical state is excluded by the core definitions  
17 underlying interventionism.

18 Classical causal exclusion arguments (e.g., Kim, 2005) are often used as arguments  
19 against non-reductive physicalism, which, as any version of physicalism, crucially relies on  
20 the assumption of the causal closure of the physical domain that prominently figures as a  
21 key premise in these arguments. In a nutshell, causal closure says that for any physical  
22 event there is a sufficient physical cause. The interventionist causal exclusion argument we  
23 will discuss is stronger because it concludes the causal inefficacy of the mental even without

---

<sup>3</sup> See, for example, (de Grefte & Gebharter, 2021; Gebharter, Graemer, & Scheffels, 2019).

1 assuming that the physical domain is causally closed. Thus, it threatens not only non-  
2 reductive physicalism but any philosophical view that assumes mental causation as well as  
3 that the mental is non-identical to the physical and supervenes on the physical, regardless  
4 of whether said view is a variant of physicalism or dualism. Therefore, we frame the main  
5 question to be pursued in this paper as whether and how mental causation is possible from  
6 an interventionist perspective given that the mental non-trivially supervenes on the physical  
7 and not to the background of the non-reductive physicalism vs. reductive physicalism  
8 debate. We believe that we are in good company in doing so and point the reader to other  
9 papers on interventionist causal exclusion which are mainly framed in terms of mental  
10 causation and epiphenomenalism rather than non-reductive vs. reductive physicalism (e.g.,  
11 Eronen & Brooks, 2014; Kistler, 2017).

12 In this paper, we zoom in on one specific assumption required to get the interventionist  
13 exclusion argument going: The mental supervenes on the physical. Supervenience is  
14 typically understood as a strict relation saying that there is no difference in higher-level  
15 supervenient properties without a difference in lower-level subvenient properties. This  
16 relation allows non-reductionists to assert the ontological non-identity of the mental and the  
17 physical. We argue that strict supervenience can be replaced by probabilistic supervenience  
18 in such a way that the interventionist exclusion argument is blocked.<sup>4</sup> Since probabilistic  
19 supervenience cannot be ruled out metaphysically, it provides an in-principle rebuff to the  
20 exclusion argument. We argue that weakening supervenience in the way we suggest does  
21 no harm because it satisfies the same core intuitions underlying the strict standard version  
22 of supervenience to which the modern scientific picture of the world typically adheres. We

---

<sup>4</sup> This is a type (B) strategy. It attacks the validity of the argument. In particular, we show that the argument's original conclusion does no longer follow from the premises given a novel and weaker interpretation of the supervenience assumption.

1 also discuss further possible objections against weakening supervenience and provide  
2 independent motivation for this move. This independent motivation suggests that  
3 probabilistic supervenience might not only block the interventionist exclusion argument but  
4 also open up a whole new area for supporters of mental causation to explore. For example,  
5 it seems to allow for a difference between mental and physical causation and to ground the  
6 intuition behind the explanatory gap argument.

7 The paper is structured as follows. In section 2, we introduce the basics of interventionism  
8 required for the rest of this paper. In section 3, we develop what we consider to be the  
9 strongest version of the interventionist exclusion argument currently on the market. In  
10 section 4, we propose to weaken supervenience and argue that this move allows the  
11 interventionist to block causal exclusion. In section 5, we finally discuss possible objections  
12 to the proposed weakening of supervenience and provide independent motivation for  
13 probabilistic supervenience. We conclude in section 6.

## 16 **2 Interventionism**

17 The basic intuition underlying interventionism is that whenever one variable  $C$  is causally  
18 relevant for another variable  $E$ , it is in principle possible to change  $E$ 's value by manipulating  
19  $C$  based on suitable interventions.<sup>5</sup> Interventions do not need to be realizable by human  
20 actions. Thus, interventionism is a non-anthropocentric manipulability theory of causation.  
21 For our endeavor, it suffices to introduce only a few basic concepts, the first of them being  
22 the notion of direct causation (Woodward, 2003, p. 55):

---

<sup>5</sup> For technical reasons, interventionism first and foremost understands causation as a relation that holds between variables. It can, thus, be considered as an account of type causation. By adding further definitions to the theory's core, one can expand interventionism to also cover token or actual causation (see, e.g., Woodward, 2003).

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23

**(DC)**  $C$  is a direct cause of  $E$  w.r.t. a set of variables  $\mathbf{V}$  if and only if there is a possible intervention on  $C$  w.r.t.  $E$  that changes  $E$ 's value while the values of all other variables in  $\mathbf{V}$  are held fixed by additional interventions.

Direct causal relationships are represented by directed edges in a causal graph.  $C \rightarrow E$ , for example, stands for  $C$  being a direct cause of  $E$ . Concatenations of such causal relations are called causal paths. A causal path of the form  $C \rightarrow \dots \rightarrow E$  is called a directed causal path from  $C$  to  $E$ . Structures consisting only of variables and causal arrows are called causal graphs. If they do not feature causal loops, they are called directed acyclic graphs or DAGs for short. In this paper, we will only consider causal structures that are DAGs.

Based on the notion of direct causation and the notion of a directed causal path, one can define the following notion of contributing causation (Woodward, 2003, p. 59):

**(CC)**  $C$  is a contributing cause of  $E$  w.r.t. a set of variables  $\mathbf{V}$  if and only if there is a directed causal path  $\pi$  from  $C$  to  $E$  and a possible intervention on  $C$  w.r.t.  $E$  that changes  $E$ 's value while the values of all other variables in  $\mathbf{V}$  not lying on  $\pi$  are held fixed by additional interventions.

This paper is not the place for a detailed motivation of **(DC)** and **(CC)**. We would like to point interested readers to (Woodward, 2003, sec. 2.3) for further details.

Note that **(DC)** and **(CC)** are both relativized to a set of variables  $\mathbf{V}$ . It can happen that a variable  $C$  turns out to be a direct or contributing cause of another variable  $E$  w.r.t. one

1 variable set  $\mathbf{V}$ , but not w.r.t. another variable set  $\mathbf{V}'$ . This can, for example, happen if there  
2 are several directed causal paths from  $C$  to  $E$  that cancel each other out. If the variables  
3 lying on these paths are represented in  $\mathbf{V}$ , but not in  $\mathbf{V}'$ , then **(DC)** and **(CC)** will be able to  
4 account for  $C$ 's causal relevance for  $E$  w.r.t.  $\mathbf{V}$ , but not w.r.t.  $\mathbf{V}'$ . This means that whether a  
5 causal relation holds according to **(DC)** and **(CC)** depends on the choice of variables. Since  
6 this paper is about metaphysics and not methodology, we also need a third causal notion  
7 that is not relativized to a particular set of variables (Woodward, 2008, p. 209):

8

9 **(CS)**  $C$  is a cause of  $E$  (simpliciter) if and only if there is an admissible variable  
10 set  $\mathbf{V}$  such that  $C$  is a contributing cause of  $E$  w.r.t.  $\mathbf{V}$ .

11

12 By an admissible variable set we mean a variable set that satisfies the following condition  
13 of independent fixability (Woodward, 2015, p. 316):

14

15 **(IF)**  $\mathbf{V}$  satisfies independent fixability if and only if any combination of values of  
16 variables in  $\mathbf{V}$  can be brought about by possible interventions.

17

18 **(IF)** can be expected to regularly be violated in the presence of non-causal relations,<sup>6</sup> but  
19 is assumed to hold for any set containing only variables describing distinct events. More  
20 precisely, distinctiveness means that for any two variables in such sets, the instantiations of  
21 the event-types represented by the values of these variables do not spatiotemporally overlap  
22 (cf. Kim, 1973; Lewis, 1986; Woodward, 2016). In order to guarantee that any causal relation

---

<sup>6</sup> For several cases where **(IF)** is satisfied even in the presence of non-causally dependent variables see, for example, (Hoffmann-Kolss, 2022).

1 output by interventionism does indeed not conflate causation with some other form of  
2 dependence, we endorse this assumption throughout the argument: If  $C$  is a cause of  $E$   
3 (simpliciter), then there is a set of variables describing distinct events such that  $C$  is a  
4 contributing cause of  $E$  w.r.t. that set of variables. We thus follow Woodward's (2015)  
5 suggestion that only sets whose variables describe distinct events and can be controlled  
6 independently should be used for evaluating causal claims.

7 Finally, we need to clarify what it means to intervene on a variable  $C$  w.r.t. another  
8 variable  $E$ . Only interventions that satisfy certain conditions will be suitable to account for  
9 causal relations. These conditions are summarized in the following definition of an  
10 intervention variable (based on Woodward, 2003, p. 98):

11

12 **(IV)**  $I_C$  is an intervention variable for  $C$  w.r.t.  $E$  if and only if

13 (i)  $I_C$  is a cause of  $C$ , and

14 (ii) if  $I_C$  is a cause of  $E$ , then only via a path going through  $C$ , and

15 (iii)  $I_C$  is probabilistically independent of any cause  $X$  of  $E$  that causes  $E$  over  
16 a path not going through  $C$ .

17

18 The notion of causation used in **(IV)** should be understood as causation (simpliciter) as  
19 defined in **(CS)**. Thus, also **(IV)** is not relativized to a particular set of variables (Woodward,  
20 2008).<sup>7</sup> Also note that in Woodward's (2003) original version there is a fourth condition  
21 requiring that the intervention variable, if active, decouples  $C$  from all its other causes. Since

---

<sup>7</sup> **(IV)** is not relativized to a particular variable set because otherwise interventionism would lead to obviously false causal claims. For details, see (Baumgartner, 2013).

1 this condition is not required to get the interventionist machinery working (Baumgartner &  
2 Drouet, 2013, pp. 186f), we can ignore it in this paper.<sup>8</sup>

3 Finally, we can define an intervention on  $C$  w.r.t.  $E$  as an intervention variable  $I_C$  for  $C$   
4 w.r.t.  $E$  taking one of its values  $i_C$  such that setting  $I_C$  to this value  $i_C$  is associated with a  
5 change of  $C$ 's value.

6 Before we proceed by presenting the interventionist exclusion argument in section 3, let  
7 us introduce a slight modification of **(IV)**. In fact, conditions (ii) and (iii) can be weakened.  
8 The original versions are intended to guarantee that any change in  $E$  associated with an  
9 intervention  $I_C = i_C$  indicates a causal influence of  $C$  on  $E$ . This can, however, be  
10 guaranteed even if  $I_C$  causes  $E$  also over a path not going through  $C$  (which violates  
11 condition (ii)) and if  $I_C$  is probabilistically dependent on another cause  $X$  of  $E$  not lying on a  
12 directed path from  $I_C$  to  $E$  going through  $C$ . This is the case if  $I_C$ 's taking on its value  $i_C$  is  
13 not associated with any change of values of any off-path variable  $X$  that causes  $E$ , i.e., any  
14 variable  $X$  that causes  $E$  not lying on a directed path from  $I_C$  to  $E$  going through  $C$ . To see  
15 the interventionist causal exclusion argument's full power<sup>9</sup>, we use this weakened version  
16 of an intervention variable throughout the paper:

- 17  
18 **(IV)\***  $I_C$  is an intervention variable for  $C$  w.r.t.  $E$  if and only if  
19 (i)  $I_C$  is a cause of  $C$ , and

---

<sup>8</sup> Our condition (i) corresponds to condition I1, (ii) to I3, and (iii) to I4 in (Woodward, 2003, p. 98).

<sup>9</sup> As we will see in the next section, the interventionist causal exclusion argument derives its strength from the result that one cannot consistently assume the existence of an intervention variable for any mental variable. Showing that there is not even a weaker intervention variable in the sense of **(IV)\*** thus makes the argument even stronger.

1 (ii) if  $I_C$  takes one of its on-values<sup>10</sup>  $i_C$ , then  $I_C = i_C$  is not associated with any  
2 change of values of any cause  $X$  of  $E$  that causes  $E$  over a path not going  
3 through  $C$ .

4  
5 We have now all the interventionist basics we need to understand the interventionist  
6 causal exclusion argument to be discussed in this paper. We introduce this argument in the  
7 next section.

### 10 **3 An interventionist version of the exclusion argument**

11 In this section we introduce what we consider to be the most threatening objection to  
12 interventionist mental causation currently on the market. The interventionist exclusion  
13 argument we focus on is based on the arguments to be found in (Baumgartner, 2010, 2013,  
14 2018). We will ignore some details and expand upon others to some extent. In particular,  
15 our version of the argument even allows the supporter of mental causation to rely on the  
16 condition of independent fixability (**IF**) and to endorse the weaker notion of an intervention  
17 variable (**IV**)\* instead of the original (**IV**). An advantage of the interventionist exclusion  
18 argument is that it does not require all the premises of the original causal exclusion  
19 argument. In particular, neither the causal closure of the physical domain nor the assumption  
20 that there is no systematic overdetermination is needed. It requires only that the mental is  
21 ontologically non-identical to the physical and that it supervenes on the physical. Thus, it not  
22 only threatens non-reductive physicalism, but any position relying on non-trivial  
23 supervenience of the mental on the physical that wants to uphold mental causation.

---

<sup>10</sup>  $I_C$ 's on-values are those values that fix  $C$  to a certain value.

1 The argument comes in the form of a *reductio ad absurdum*. It runs as follows: Assume  
2 that the mental variable  $M$  is a cause (simpliciter) of another causal variable  $E$ . This is the  
3 assumption that is challenged by the argument by showing that it leads to inconsistencies.  
4 Let us further assume that the mental supervenes on the physical, meaning that there will  
5 also be a variable  $P$  such that any change in  $M$ -values is associated with a change in  $P$ -  
6 values. From  $M$  being a cause (simpliciter) of  $P$  it follows with **(CS)** that there is an  
7 admissible set of variables  $\mathbf{V}$  containing  $M$  and  $E$  such that  $M$  turns out to be a contributing  
8 cause of  $E$  w.r.t.  $\mathbf{V}$ . This set will only contain variables representing distinct event types and,  
9 thus, it will also satisfy **(IF)**, meaning that all the variables in  $\mathbf{V}$  can be independently fixed  
10 by interventions. Note that it follows from this that  $\mathbf{V}$  does not contain  $M$ 's supervenience  
11 base  $P$  since due to supervenience not all combinations of  $M$ -values and  $P$ -values describe  
12 distinct events and can be brought about independently by interventions. Now for  $M$  to be  
13 a contributing cause of  $E$  w.r.t.  $\mathbf{V}$  there must, according to **(CC)**, exist an intervention  $I_M =$   
14  $i_M$  on  $M$  w.r.t.  $E$  that is associated with a change in  $E$  if the values of all off-path variables  
15 are fixed by additional interventions. So far so good.

16 Let us turn to  $M$ 's supervenience base  $P$  now. Since  $\mathbf{V}$  is an admissible variable set, also  
17 the set  $\mathbf{V}'$  that results from  $\mathbf{V}$  by replacing  $M$  by  $P$  will be admissible. The reason for this is  
18 that  $\mathbf{V}'$  will not contain any pair of variables whose values represent event-types with  
19 spatiotemporally overlapping instantiations because also  $\mathbf{V}$  did not contain such a pair of  
20 variables. To see why this is the case, let us assume that  $\mathbf{V}'$  would indeed contain two  
21 variables with values representing spatiotemporally overlapping event-types. These two  
22 variables are either both different from  $P$  or they are not. If they are both different from  $P$ ,

1 then also  $\mathbf{V}$  would contain them, which is excluded by  $\mathbf{V}$  being an admissible variable set.  
2 Thus, one of these variables needs to be  $P$ . (Let us label the other one  $X$ .) If this would be  
3 the case, then also  $M$  would feature values representing event-types that spatiotemporally  
4 overlap with event-types represented by values of  $X$ . This follows from the values of  $P$  being  
5 the realizers of the values of  $M$  and the fact that realizers cannot spatiotemporally exceed  
6 what they realize. But also  $M$  featuring values representing event-types that  
7 spatiotemporally overlap with event-types represented by values of  $X$  is already excluded  
8 by  $\mathbf{V}$  being admissible.

9 Since we already established that there exists an intervention  $I_M = i_M$  on  $M$  w.r.t.  $E$  that  
10 is associated with a change in  $E$  when the values of all off-path variables in  $\mathbf{V}$  are fixed by  
11 additional interventions and since  $M$  supervenes on  $P$ , we can infer that there also exists  
12 an intervention  $I_P = i_P$  on  $P$  w.r.t.  $E$  that is associated with a change in  $E$  when the values  
13 of all off-path variables in  $\mathbf{V}'$  are fixed by additional interventions. The reason for this is how  
14  $P$ -values are mapped onto  $M$ -values due to supervenience. All that is needed is to be able  
15 to bring about one of the  $P$ -values  $p$  realising one of those  $M$ -values  $m$  for which we  
16 already established that  $I_M = i_M$  is associated with changes in  $E$ . Since  $\mathbf{V}'$  is an admissible  
17 variable set, it follows from **(IF)** that there will be such an intervention  $I_P = i_P$  on  $P$  w.r.t.  $E$ .  
18 This intervention will, since it sets  $M$  to  $m$  due to the fact that  $m$  is realized by  $p$ , also be  
19 associated with changes in  $E$ . It then follows from **(CC)** that  $P$  is a contributing cause of  $E$   
20 w.r.t.  $\mathbf{V}'$  and from this with **(CS)** that also  $P$  is a cause (simpliciter) of  $E$ .

21 We have found so far that both  $M$  and  $P$  are causes (simpliciter) of  $E$ . Let us now focus  
22 on the intervention variable  $I_M$ . From **(IV)\*** it follows that  $I_M$  is a cause (simpliciter) of  $M$ .

1 Thus, from **(CS)** it follows that there must be an admissible variable set  $\mathbf{V}''$  containing  $I_M$   
2 and  $M$  such that  $I_M$  is a contributing cause of  $M$  w.r.t.  $\mathbf{V}''$ . From **(CC)** it then follows that  
3 there exists an intervention  $I_{I_M} = i_{I_M}$  on  $I_M$  w.r.t.  $M$  such that  $I_{I_M} = i_{I_M}$  is associated with  
4 a change in  $M$  when the values of all off-path variables in  $\mathbf{V}''$  are fixed by additional  
5 interventions. Let  $\mathbf{V}'''$  be the variable set we get from  $\mathbf{V}''$  by replacing  $M$  by  $P$ . Again, since  
6  $\mathbf{V}''$  was admissible, also  $\mathbf{V}'''$  will be admissible (for the very same reasons for which  $\mathbf{V}'$   
7 turned out as admissible if  $\mathbf{V}$  is). And since  $M$  supervenes on  $P$  and  $I_{I_M} = i_{I_M}$  is associated  
8 with a change in  $M$  when the values of all off-path variables in  $\mathbf{V}''$  are fixed by additional  
9 interventions, it will also be associated with a change in  $P$  when the values of all off-path  
10 variables in  $\mathbf{V}'''$  are fixed by additional interventions. The reason for this is that due to  
11 supervenience no two different  $M$ -values can be realised by one and the same  $P$ -value. So,  
12 for any change in  $M$ -values due to  $I_{I_M} = i_{I_M}$  there must be a corresponding change in  $P$ -  
13 values. It then follows with **(CC)** that  $I_M$  is a contributing cause of  $P$  w.r.t.  $\mathbf{V}'''$  and from the  
14 latter with **(CS)** that  $I_M$  is a cause (simpliciter) of  $P$ .

15 Here comes the problem: Since any directed causal path between  $M$  and  $P$  is excluded  
16 by the typical assumption that supervenience is a non-causal relation<sup>11</sup>, it follows that  $I_M$   
17 causes  $E$  over two different causal paths, one going through  $M$  and one going through  $P$   
18 such that any intervention  $I_M = i_m$  that is associated with changes in  $M$ -values is, at the  
19 same time, associated with changes in  $P$ -values. This, however, contradicts **(IV)\***(ii) and,  
20 thus, the earlier result that  $I_M$  is an intervention variable for  $M$  w.r.t.  $E$ . Since nothing hinged

---

<sup>11</sup> For a recent proposal to understand supervenience as a causal relation, see (Leuridan, 2012; Leuridan & Lodewyckx, 2020). We point the interested reader to these papers but do not explore the strategy discussed in them further here.

1 on the particular choice of the variables we used, this result generalizes: From the  
2 assumption that any mental variable  $M$  is a cause (simpliciter) of another variable  $E$  it  
3 follows that there exists an intervention variable  $I_M$  for  $M$  w.r.t.  $E$  and, at the same time,  
4 that no such intervention variable exists. Thus, by *reductio* it follows that no mental variable  
5 can be a cause (simpliciter) of any other variable whatsoever.

6 Let us briefly add two comments about this argument: The first one is that the argument  
7 works even under the assumption that all the variable sets used in the argument satisfy  
8 independent fixability **(IF)**. This is remarkable because several authors (most prominently  
9 Woodward, 2015) have argued that interventionist exclusion arguments can be blocked by  
10 committing oneself to only using variable sets satisfying **(IF)** for causal evaluations along  
11 the lines of **(DC)** and **(CC)**. This is exactly what we did in our version of the interventionist  
12 exclusion argument as presented above. We assumed **(IF)** whenever we applied **(DC)** or  
13 **(CC)** and only required to hold fixed variables in those specific sets. But as it turned out, not  
14 even restricting oneself to sets satisfying **(IF)** does in the end improve the situation for the  
15 supporter of mental causation.

16 The second comment is that Woodward (2015) suggested that the argument (or a version  
17 of it) could be blocked by allowing for interventions that violate condition **(IV)\***(ii) in cases  
18 where supervenience relationships are involved. For our argument this would mean that the  
19 consequence that  $I_M$  is an intervention variable that influences  $M$  as well as  $P$  at the same  
20 time but over two different causal paths would not amount to a contradiction anymore.  
21 However, Baumgartner (2018) showed that in such cases one cannot decide whether the  
22 causal work that brings about changes in  $E$  is done by the directed path going through  $M$ ,  
23 by the directed path going through  $P$ , or by both paths together. Thus,  $M$ 's causal efficacy  
24 cannot be established conclusively and the threat of epiphenomenalism would still stand.

1 Since the modification to the strict notion of supervenience used in both arguments that we  
2 will propose in the next section already blocks the first argument, there is no need to further  
3 modify the notion of an intervention variable **(IV)\*** and we will stick with the version of the  
4 interventionist exclusion argument developed in some detail above.

5

6

#### 7 **4 Probabilistic supervenience to the rescue**

8 In this section we propose to replace the strict notion of supervenience used in the  
9 interventionist exclusion argument introduced in section 3 by a weaker probabilistic version.  
10 After that, we show how this probabilistic version of supervenience can be used to block the  
11 interventionist exclusion argument.

12 To make our overall strategy more transparent, it is useful to start with a more precise  
13 definition of strict supervenience:<sup>12</sup>

14

15 **(SUP)** *M* supervenes on *P* if and only if (i) each *M*-value is realized by some *P*-  
16 value, and (ii) *M*'s value does not vary anymore once *P*'s value is fixed.

17

18 Note that **(SUP)** is stated for variables, which makes it directly applicable to the  
19 background of an interventionist understanding of causation. (i) guarantees that no *M*-value  
20 can occur without a corresponding *P*-value being instantiated at the same time. (ii) reflects

---

<sup>12</sup> Prominent definitions of supervenience to be found in the literature often feature necessity operators. (For an overview see, e.g., McLaughlin & Bennett, 2021). These stronger notions imply **(SUP)**. Running the interventionist causal exclusion argument with the weaker version **(SUP)**, as we did in sec. 3, thus reveals its full strength.

1 the supervenience assumption: Whenever  $M$  changes its value, then also  $P$  does, or the  
2 other way round: If  $P$ 's value is fixed, then also  $M$ 's value is.

3 What caused the problem for mental causation in section 3 was that any cause of  $M$   
4 turned out to be a common cause of  $M$  and  $P$ . This is a direct consequence of **(SUP)**. So,  
5 what we need to allow for in order to block the interventionist exclusion argument is a notion  
6 of supervenience that avoids this consequence. Before we propose such a notion, it will be  
7 useful to state **(SUP)** in probabilistic terms:

8

9 **(SUP)\***  $M$  supervenes on  $P$  if and only if (i) each  $M$ -value is realized by some  $P$ -  
10 value, and (ii) for all  $P$ -values  $p$  there is an  $M$ -value  $m$  such that  
11  $Pr(m|p) = 1$ .

12

13 Now it becomes easy to see how strict supervenience can be weakened. Instead of  
14 requiring that every  $P$ -value  $p$  determines  $M$  to take one of its values  $m$  with probability  
15 1, we only demand that every change in  $M$ -values makes a probabilistic difference for at  
16 least one of  $P$ 's values. We thus arrive at the following notion of probabilistic  
17 supervenience:

18

19 **(SUP)\*\***  $M$  supervenes on  $P$  if and only if (i) each  $M$ -value is realized by some  $P$ -  
20 value, and (ii) for all  $M$ -values  $m$  and  $m'$  (with  $m \neq m'$ ) there is a  $P$ -value  
21  $p$  such that  $Pr(p|m) \neq Pr(p|m')$ .

22

1 Note that the move from strict to probabilistic supervenience preserves important  
2 intuitions about supervenience: Supervenience should anchor or ground the mental in the  
3 ongoing in the physical. In other words, supervenience is assumed to restrict what is  
4 possible on higher levels given what is happening on the fundamental physical level. This is  
5 what makes the supervenience relation so attractive to non-reductionists about mental  
6 causation: It allows for the non-identity of physical and mental properties without claiming  
7 that mental properties can exist without some fundamental physical properties underlying  
8 them.<sup>13</sup> Like strict supervenience, probabilistic supervenience can do the same job,  
9 depending on the specific probabilistic pattern between  $M$ -values and  $P$ -values. Once the  
10 physical variable  $P$  on which the mental variable  $M$  supervenes is fixed to a certain value  
11  $p$ , some mental states might be compatible with  $p$ , while others might not be compatible.<sup>14</sup>  
12 The difference to strict supervenience is that its probabilistic cousin is less restrictive. Now  
13 it can happen that more than one mental state  $m$  is compatible with a physical state  $p$ . How  
14 restrictive a particular supervenience relation is, in the end, an empirical question that  
15 cannot be answered a priori. Note, however, that also strict supervenience is not inferred  
16 from empirical data showing how mental properties actually correlate with physical  
17 properties, but is rather accepted on purely metaphysical grounds. If probabilistic  
18 supervenience does the same job without facing some of the problems of strict  
19 supervenience, this constitutes a good reason to prefer it. In section 5 we will consider

---

<sup>13</sup> Craver (2017) discusses the same metaphysical relation under the label “stochastic physical supervenience”. He concludes that while such a relation is conceptually possible, it is neither well-motivated nor helpful, although he sees some potential of this relation in addressing the causal exclusion argument. We argue that one should consider probabilistic supervenience because it not only blocks the interventionist exclusion argument, but also offers other interesting empirical and metaphysical possibilities.

<sup>14</sup> Note that  $Pr(p|m) \neq Pr(p|m')$  in **(SUP)\*\*** implies that  $P$  probabilistically depends on  $M$ . Since probabilistic dependence is a symmetric relation, it thus further implies that also  $M$  probabilistically depends on  $P$ . Hence, the supervenience base will always have some probabilistic influence on the supervening variable.

1 different arguments in favor of probabilistic supervenience, as well as some conceptual  
2 consequences of accepting it.

3 Let us now come back to the question of how **(SUP)\*\*** can help us to overcome the  
4 interventionist exclusion argument. To this end, we provide a simple model illustrating how  
5  $M$ 's value can be changed by an intervention variable  $I_M$  for  $M$  w.r.t.  $E$  without changing  
6  $P$ 's value at the same time. Assume  $M$  has two values  $m_1, m_2$  and its supervenience base  
7  $P$  has four values  $p_1, p_2, p_3, p_4$ . Further assume that the following conditional probabilities  
8 characterize the (probabilistic) supervenience relationship between  $M$  and  $P$ :

$$Pr(m_1|p_1) = 0.75$$

$$Pr(m_1|p_2) = 0.5$$

$$Pr(m_1|p_3) = 0.25$$

$$Pr(m_1|p_4) = 0$$

10  
11 According to this characterization,  $m_1$  can be realised by  $p_1, p_2$ , and  $p_3$ , while  $m_2$  can  
12 be realised by all four  $P$ -values. In particular, the probabilities that each of the individual  $P$ -  
13 values realizes  $m_1$  and  $m_2$ , were  $m_1$  and  $m_2$  be instantiated, are as follows:<sup>15</sup>

$$Pr(p_1|m_1) = 0.50 \quad Pr(p_1|m_2) = 0.1$$

---

<sup>15</sup> For the sake of simplicity, we assume that  $Pr(p_i) = 0.25$  for  $i \in \{1,2,3,4\}$ . The conditional probabilities  $Pr(p_i|m_j)$  with  $j \in \{1,2\}$  can be computed with Bayes' theorem as  $Pr(p_i|m_j) = \frac{Pr(m_j|p_i)Pr(p_i)}{\sum_{p_i, m_j} Pr(m_j|p_i)Pr(p_i)}$

$$Pr(p_2|m_1) = 0.33$$

$$Pr(p_2|m_2) = 0.2$$

$$Pr(p_3|m_1) = 0.16$$

$$Pr(p_3|m_2) = 0.3$$

$$Pr(p_4|m_1) = 0.00$$

$$Pr(p_4|m_2) = 0.4$$

1

2 Now the interventionist exclusion argument can be blocked as follows: We proceed  
3 exactly as in the argument outlined in section 3 until we have established that both the  
4 mental variable  $M$  and the physical variable  $P$  are causes (simpliciter) of the causal variable  
5  $E$ . The difference is how we proceed from there. In the original argument, we were able to  
6 show that  $I_M$  is not only a cause simpliciter of  $M$ , but also of  $P$  and, more importantly, that  
7 any change brought about in  $M$ -values due to  $I_M = i_M$  corresponded to a change in  $P$ -  
8 values since no two different  $M$ -values can be realized by one and the same  $P$ -value due  
9 to strict supervenience. But this is not the case if we replace strict by probabilistic  
10 supervenience. If we use an intervention  $I_M = i_M$  to change  $M$ 's value from  $m_1$  to  $m_2$ , for  
11 example, then  $I_M = i_M$  does not necessarily also lead to a change in  $P$ -values. The reason  
12 for this is that both  $m_1$  and  $m_2$  are compatible with the physical realisers  $p_1$ ,  $p_2$ , and  $p_3$ .  
13 Thus, it becomes conceptually possible that there exists an intervention  $I_M = i_M$  such that  
14 it changes  $M$ 's value while it does not change  $P$ 's value. This, in turn, means that it also  
15 becomes possible to consistently establish that a mental variable  $M$  is a cause simpliciter  
16 of a causal variable  $E$ , which shows that the interventionist exclusion argument can be  
17 blocked.

18

19

## 1 **5 Independent motivation and possible objections**

2 Showing that replacing strict by probabilistic supervenience can block the interventionist  
3 exclusion argument is one thing. Arguing that this move is reasonable and that assuming  
4 probabilistic instead of strict supervenience is well motivated is another. In this section, we  
5 provide some independent motivation for probabilistic supervenience and try to dispel  
6 possible worries.

### 9 **5.1 Supervenience in the image of causation**

10 We would like to put forward the following independent motivation for probabilistic  
11 supervenience. It consists in a brief comparison with causation, historically as well as  
12 conceptually. Historically, causation was often viewed as the glue that holds reality together.  
13 This view can be traced back to Aristotle (Physics II 7; see also Stein, 2012). Why do  
14 some types of events but not others regularly succeed each other in time? How can we  
15 understand these patterns though these types of events are not connected conceptually?  
16 For example, copper expands when heated, but the concepts of copper and heating do not  
17 imply this behavior in any way. The answer is to assume a contingent causal relation that  
18 can do that job. Learning about causal relations provides us with information about which  
19 types of events do and do not succeed a certain type of event in time and often even with  
20 information about the probability with which the effect can be expected to succeed the  
21 cause. Vice versa, knowing the underlying causal relation, observing a phenomenon allows  
22 us to say which events might and which ones might not have preceded the effect. Thus,  
23 causation structures the space of logically possible successions of types of events in both  
24 directions, towards the future and the past.

1 From a formal point of view, supervenience plays a similar role, but vertically rather than  
2 horizontally. While causation restricts the space of events that could precede or succeed an  
3 event of a certain type, supervenience restricts the space of possible realizers of a state on  
4 a lower level as well as the space of states that might emerge from said state (and possibly  
5 other states) at a higher level. Like in the case of causation, what realizes what and what  
6 emerges from what are not questions that can be answered a priori based on conceptual  
7 analysis. These are empirical matters (cf. Leuridan, 2017). Actual research is required, for  
8 example, to find out that water is H<sub>2</sub>O or to determine which brain processes can constitute  
9 which mental states. Summarizing, we can say that while causation anchors or grounds later  
10 events in earlier events, supervenience anchors or grounds higher-level states in lower-level  
11 states.<sup>16</sup>

12 If we now take another look at the history of causation and especially at the prevalent  
13 view on causation up until the 18<sup>th</sup> century – as held, for example, by Aristotle, Descartes,  
14 Spinoza, and Newton – we can observe that causation was typically understood as a strict  
15 relation. This relation was similar to what we would now call *metaphysical necessitation*,  
16 even though these particular terms were not employed. The idea was that causes fully  
17 determine their effects and that this is the only way how causal relations can hold reality  
18 together<sup>17</sup>. More recent developments in philosophy, however, showed that there are many

---

<sup>16</sup> For a more formal argument for the claim that supervenience shares important features with causation, see (Gebharder, 2017a).

<sup>17</sup> Descartes famously diverged from Aristotle regarding the number of types of causes. He posited only one – the efficient cause – as opposed to Aristotle's four. Despite this difference, both Descartes and Aristotle agreed that causation is a strict relation and that causes necessitate their effects: Once the efficient cause is in place, its effect must follow (Schmaltz, 2008; Stein, 2012). This view was accepted by post-Cartesian philosophers and scientists until Hume, who was the first to question the received view. Whether Hume denied that the relation of causality itself is a metaphysical necessity relation, or whether he only denied that we can ever be justified in making particular causal claims (which, according to him, were indeed claims about the necessitation relation between the cause and the effect), is a matter of dispute (cf. Beebe, 2011). However, Hume started the progressive weakening of the modality of causal relations, paving the way for today's probabilistic approaches.

1 causal relations out there that are at best probabilistic. Nowadays there is a whole plethora  
2 of probabilistic theories or at least probability-friendly accounts of causation on the market  
3 (see, e.g., Cartwright, 1979; Eells, 1991; Pearl, 2000; Spirtes, Glymour, and Scheines, 1993;  
4 Suppes, 1970) and many philosophers do not consider it as problematic to accept that  
5 causation might be an inherently probabilistic relation.

6 We propose that it might be time for a similar step forward when it comes to the concept  
7 of supervenience. Maybe also strict supervenience is an artefact of a bygone age and it is  
8 time now to embrace supervenience in its full richness instead of only focusing on the  
9 extreme case. Note that probabilistic supervenience, like probabilistic causation, is still well  
10 capable of performing the job outlined above: It still restricts what is possible at higher as  
11 well as at lower levels given a certain event or phenomenon. Downwards looking, even strict  
12 supervenience allows for more than one lower-level realizer of a given phenomenon due to  
13 multiple realizability. Maybe it is time to allow for the same one-to-many relation when  
14 looking upwards as well. This could still constrain the space of logically possible higher-level  
15 states that might emerge from a given lower-level state, the only difference being that it  
16 would in principle allow for more than one such state.

## 19 **5.2 Probabilistic supervenience vs. causal closure**

20 In this subsection we want to discuss a possible conceptual worry one might have about  
21 probabilistic supervenience. Assume that we find a case in which it turns out that, according  
22 to interventionism, a mental variable  $M$  is causally efficacious w.r.t. a causal variable  $E$ .  
23 Now one might worry that this finding would stand in contradiction to another typical  
24 naturalistic assumption: the closure of the physical domain. Physical closure says that any  
25 physical state can be sufficiently explained by pointing to its physical causes. Citing a non-

1 physical cause—that would be *M* in our example—cannot add anything to whether such a  
2 state is instantiated. For the moment, let *E* model the physical state to be explained. In terms  
3 of interventionism this would mean that once the value of *P*—let us assume that *P* is *E*'s  
4 sufficient physical cause—is fixed, intervening on *M* cannot have any influence on *E*  
5 anymore. This is true if we assume strict supervenience since, as we saw earlier, *M*'s value  
6 cannot be changed once *P*'s value is fixed. However, the same does not hold if we assume  
7 probabilistic supervenience. Assuming the latter allows in principle for an intervention on *M*  
8 w.r.t. *E* that is associated with a change of *E*'s value even if *P*'s value does not change. But  
9 then, obviously, *M* can have a causal influence on *E* over and above its sufficient physical  
10 cause *P* and, therefore, causal closure would be violated.

11 We believe that nothing is wrong with the above line of reasoning. Friends of probabilistic  
12 supervenience do, however, have several options to respond. The first kind of response  
13 they could give is to simply bite the bullet. If the physical domain is causally closed, then the  
14 mental cannot add anything and vice versa: If the mental causally contributes to what is  
15 going on at the physical level, then causal closure cannot be upheld. As a metaphysician,  
16 one needs to make a choice here. Either go for causal closure and throw mental to physical  
17 causation out the window, or the other way round. It is not that unusual that supporters of  
18 mental causation decide in favor of the latter option, may that be more explicitly or rather  
19 implicitly. Take List and Menzies' (2009) argument in favor of mental causation as an  
20 example. In their account, an event's mental causes can even exclude its physical causes  
21 altogether.

22 One could also give a somewhat more nuanced answer when deciding in favor of the  
23 closure of the physical domain. One might drop mental to physical causation but keep

1 mental causation in the form of mental to mental causation. In terms of interventionism, one  
2 could formulate this as follows: Because the physical domain is causally closed, intervening  
3 on  $M$  will not lead to any change in  $E$  if  $E$  stands for a physical state. However, if  $E$   
4 represents a mental state, then intervening on  $M$  might well lead to a change in  $E$  even if  
5  $M$ 's supervenience base  $P$  does not change. Nothing we assumed so far excludes that such  
6 a change in  $E$  associated with an intervention on  $M$  must change  $E$ 's physical  
7 supervenience base  $P^*$  as well. But if this is possible and we can change the mental variable  
8  $E$ 's value by means of intervening on  $M$  without adding anything to  $P$ 's causal effect on  $P^*$ ,  
9 then physical closure is not violated at all. Keep in mind that if we go for this route, the mental  
10 is still sufficiently anchored or grounded in the physical. It still does not follow that anything  
11 goes at the mental level; the mental variable  $E$ 's values can still be sufficiently constrained  
12 by  $E$ 's supervenience base  $P^*$ .

### 15 **5.3 Probabilistic supervenience and physicalism**

16 Finally, let us discuss another possible worry one might have about probabilistic  
17 supervenience.<sup>18</sup> Typically, causal exclusion arguments are launched based on the core  
18 assumptions of non-reductive physicalism and are aimed at showing that non-reductive  
19 physicalism is untenable and should be discarded in favor of reductive physicalism. Now  
20 one might worry that replacing strict supervenience by probabilistic supervenience renders  
21 one's position a non-physicalist position right from the beginning. Thus, countering the  
22 interventionist causal exclusion argument by relying on probabilistic supervenience would

---

<sup>18</sup> We are indebted to an anonymous reviewer for pushing us to reflect more on this possible worry.

1 not help one since it would at the same time mean to give up the position one wanted to  
2 defend, viz. non-reductive physicalism.

3 To counter this worry, let us first remind the reader of what we announced in the  
4 introduction and saw in full detail in section 3: The interventionist causal exclusion argument  
5 does not require all the premises classical exclusion arguments require. In particular, it only  
6 requires the two assumptions that the mental is non-identical to the physical and that the  
7 mental (strictly) supervenes on the physical. While these assumptions are both crucial  
8 theses of non-reductive physicalism, they are not sufficient for a position to be a type of non-  
9 reductive physicalism. Non-reductive physicalism does typically also crucially depend on the  
10 thesis that the physical domain is causally closed, that every causal event has a causally  
11 sufficient physical cause or explanation. We already discussed how probabilistic  
12 supervenience relates to the causal closure of the physical domain in subsection 5.2.  
13 However, since the interventionist exclusion argument does not require that latter  
14 assumption, it does not only threaten non-reductive physicalism, but any philosophical view  
15 committed to mental causation that relies on the non-identity of the mental and the physical  
16 and the assumption that the mental supervenes on the physical, including some dualist  
17 positions. Thus, a way to block the interventionist exclusion argument is not only attractive  
18 for non-reductive physicalists.

19 Be that as it may, one might still worry how a subscriber to non-reductive physicalism can  
20 consistently uphold their position after exchanging strict supervenience for its probabilistic  
21 counterpart. One specific worry may be that probabilistic supervenience leads to violations  
22 of the causal closure of the physical domain. As we saw in subsection 5.2, probabilistic  
23 supervenience does not automatically imply this. It is conceptually compatible with all of the  
24 following possibilities: (i) Mental variables causally influence only other mental variables, (ii)  
25 mental variables causally influence only physical variables, (iii) mental variables sometimes

1 causally influence both other mental and physical variables, and (iv) mental variables  
2 causally influence no other variable whatsoever. Which of these possibilities is actually the  
3 case is a contingent matter that we cannot decide simply by endorsing probabilistic  
4 supervenience. It depends on the world, in particular on whether the interventions  
5 supporting each of these possibilities exist which, in turn, depends on the specific conditional  
6 probabilities of the mental variables' values given their physical supervenience base  
7 variables' values to be found in the world. Now possibility (i) does not violate causal closure.  
8 Thus, if the world is such that (i) comes out as true, one can clearly uphold non-reductive  
9 physicalism. Since probabilistic supervenience does not determine which one of these four  
10 possibilities comes out as true, replacing strict supervenience by probabilistic supervenience  
11 does not automatically render one's position different from non-reductive physicalism.

12 Summarizing what we found so far, non-reductive physicalism is in principle compatible  
13 with probabilistic supervenience. There is still concern that the claim that several mental  
14 states are now compatible with a single physical state automatically makes one a  
15 dualist. This would go against another crucial requirement of non-reductive physicalism:  
16 Non-reductive physicalism is typically understood as a monist position. A strategy to dispel  
17 such a worry goes as follows: One way to conceptualize physicalism is to assume that all  
18 the entities in the world are physical. Thus, physicalism is an entity monism. To acknowledge  
19 another crucial commitment of non-reductive physicalism, viz. that the mental is not identical  
20 to the physical, one can then assume that some properties are mental while others are  
21 physical, but that both are instantiated by physical entities. Thus, one ends up with a position  
22 that is monist when it comes to entities (they are all physical), but dualist when it comes to  
23 properties (they are either mental or physical). Hence, one's version of non-reductive  
24 physicalism is a monism in a relevant sense: There are only physical entities. At the same  
25 time, it is non-reductive because probabilistic supervenience provides a metaphysical

1 foundation for another intuition crucial to non-reductive physicalists: namely, the intuition  
2 that at least some mental properties, such as phenomenal experiences, cannot be  
3 reductively explained by any physical properties (Levine 1983; Kim 2005). If the relation  
4 between mental and physical properties is probabilistic rather than strict supervenience, it  
5 becomes clearer why certain mental properties, while supervening on physical properties,  
6 cannot be reductively explained: there exists a metaphysical possibility for slightly different  
7 mental properties to supervene on the same physical base.

8 Finally, we acknowledge that one might still have worries. For example, not everyone  
9 believes that even strict supervenience is a strong enough assumption for physicalism.  
10 Kroedel and Schulz (2016), for example, propose that grounding should be used in causal  
11 exclusion contexts. If one considers strict supervenience as too weak, then surely  
12 probabilistic supervenience will be even worse. These are interesting and original  
13 objections, but since they do not specifically threaten probabilistic supervenience, but  
14 supervenience in general, we believe that this paper is not the right place to discuss them  
15 in more detail. For now, we content ourselves with emphasizing that the majority view still  
16 relies on supervenience as an essential assumption for non-reductive physicalism and with  
17 pointing to our argumentation at the end of section 2 for why probabilistic supervenience still  
18 plays the same role as strict supervenience for anchoring or grounding the mental in the  
19 physical.

20  
21

## 22 **6 Conclusion**

23 We argued that previous interventionist approaches to saving mental causation from the  
24 causal exclusion argument do not work because of the assumption of strict supervenience.  
25 To this end, we put forward what we believe to be the strongest version of the interventionist

1 causal exclusion argument currently on the market. From the premises required for the  
2 classical exclusion argument it only requires the assumption that the mental non-trivially  
3 supervenes on the physical. In addition, it works with the weaker notion of an intervention  
4 variable **(IV)\*** and even under the assumption that all variable sets used for evaluating causal  
5 claims need to satisfy independent fixability **(IF)**.

6 As we saw, strict supervenience renders any intervention on any mental variable  
7 impossible. Consequently, no mental variable can be a cause of any other variable.  
8 However, if the assumption of strict supervenience is replaced by the assumption of  
9 probabilistic supervenience, the values of mental variables are not fully determined by the  
10 values of the variables representing their physical supervenience bases, which renders  
11 interventions on mental variables possible. Consequently, if probabilistic supervenience is  
12 accepted, then the mental can be causally efficacious within the interventionist account of  
13 causation. At this point we would like to emphasize that this does not yet establish that there  
14 actually is any mental causation. It only paves the road for mental causation by showing that  
15 it is compatible with the other assumptions used in the interventionist causal exclusion  
16 argument.

17 Moreover, probabilistic supervenience would not only save mental causation from the  
18 causal exclusion argument, but also has other virtues. First, it is compatible with the  
19 naturalistic worldview because it shows that mental properties are not completely  
20 disconnected from their underlying physical properties but are probabilistically anchored in  
21 them. Second, it allows for symmetric multiple realizability: Not only the same mental state  
22 can be realized by different physical states, but also different mental states can be realized  
23 by the same physical state. This provides a metaphysical foundation for qualitatively  
24 different subjective mental states given the same underlying physical states, and thereby  
25 grounds the intuition behind the explanatory gap argument (Levine 1983). Finally, it leads

1 to new conceptual developments in the debate on mental causation, because it shows that  
2 the thesis of the causal closure of the physical is compatible with the thesis of the causal  
3 efficacy of the mental only if the casual efficacy of the mental is restricted to the domain of  
4 the mental. These considerations do not conclusively show that the mental actually  
5 probabilistically supervenes on the physical. What we wanted to achieve is rather to offer a  
6 metaphysically possible and somewhat plausible in-principle way to block the interventionist  
7 causal exclusion argument. We argued that probabilistic supervenience is a possibly useful  
8 concept for supporters of mental causation and that it is not that outlandish as it might seem  
9 at first glance.

10  
11  
12 **Acknowledgments:** We would like to thank two anonymous reviewers for helpful  
13 comments on an earlier version of this paper. This work was funded by the German  
14 Research Foundation (DFG), project number 441311834 (Alexander Gebharder) and  
15 Research group Inductive Metaphysics (FOR 2495), research grant SCHU 1566/11-2 (Maria  
16 Sekatskaya).

17  
18  
19 **Conflict of interest statement**

20 The authors declare that there is no conflicts of interest.

21  
22  
23 **References**

- 1 Aristotle (1984). *Physics*. In J. Barnes (Ed.), *Complete works of Aristotle, volume 1: The*  
2 *revised Oxford translation* (pp. 315–446). Transl. by R. P. Hardie & R. K. Gaye. Princeton  
3 University Press.
- 4 Baumgartner, M. (2010). Interventionism and epiphenomenalism. *Canadian Journal of*  
5 *Philosophy*, 40(3), 359–383.
- 6 Baumgartner, M. (2013). Rendering interventionism and non-reductive physicalism  
7 compatible. *Dialectica*, 67(1), 1–27.
- 8 Baumgartner, M. (2018). The inherent empirical underdetermination of mental causation.  
9 *Australasian Journal of Philosophy*, 96(2), 335–350.
- 10 Baumgartner, M., & Drouet, I. (2013). Identifying intervention variables. *European Journal*  
11 *for Philosophy of Science*, 3(2), 183–205.
- 12 Baumgartner, M., & Gebharder, A. (2016). Constitutive relevance, mutual manipulability, and  
13 fat-handedness. *British Journal for the Philosophy of Science*, 67(3), 731–756.
- 14 Beebe, H. (2011). *Hume on causation*. London: Routledge
- 15 Cartwright, N. (1979). Causal laws and effective strategies. *Nous*, 13, 419–437.
- 16 Craver, C. (2017). Stochastic supervenience. In M. P. Adams, Z. Biener, U. Feest, & J. A.  
17 Sullivan (Eds.), *Eppur si muove: Doing history and philosophy of science with Peter*  
18 *Machamer: A collection of essays in honor of Peter Machamer* (pp. 163–170). Cham:  
19 Springer.
- 20 de Grefte, J., & Gebharder, A. (2021). The causal theory of knowledge revisited: An  
21 interventionist approach. *Ratio*, 34(3), 193–202.
- 22 Eells, E. (1991). *Probabilistic causality*. Cambridge: Cambridge University Press.
- 23 Eronen, M. & Brooks, D. (2014) Interventionism and supervenience: A new problem and  
24 provisional solution. *International Studies in the Philosophy of Science*, 28(2), 185–202.

- 1 Gebharter, A. (2017a). Causal exclusion and causal Bayes nets. *Philosophy and*  
2 *Phenomenological Research*, 95(2), 353–375.
- 3 Gebharter, A. (2017b). Causal exclusion without physical completeness and no  
4 overdetermination. *Abstracta — Linguagem, Mente E Ação*, 10, 3–14.
- 5 Gebharter, A., Graemer, D., & Scheffels, F. H. (2019). Establishing backward causation on  
6 empirical grounds: An interventionist approach. *Thought: A Journal of Philosophy*, 8(2),  
7 129–138.
- 8 Hitchcock, C. (2012). Theories of causation and the causal exclusion argument. *Journal of*  
9 *Consciousness Studies*, 19(5-6), 40–56.
- 10 Hoffmann-Kolss, V. (2014). Interventionism and higher-level causation. *International*  
11 *Studies in the Philosophy of Science*, 28(1), 49–64
- 12 Hoffmann-Kolss, V. (2021). Interventionism and non-causal dependence relations: New  
13 work for a theory of supervenience. *Australasian Journal of Philosophy*. Advance online  
14 publication. doi:10.1080/00048402.2021.1941154
- 15 Hume, D. (1738/1975). *A treatise of human nature*. Oxford: Clarendon Press.
- 16 Hume, D. (1748/1999). *An enquiry concerning human understanding*. Oxford: Oxford  
17 University Press.
- 18 Kim, J. (1973). Causes and counterfactuals. *The Journal of Philosophy*, 70, 570–72.
- 19 Kim, J. (2005). *Physicalism, or something near enough*. Princeton University Press.
- 20 Kinney, D. (2023). Bayesian networks and causal ecumenism. *Erkenntnis*, 88(1), 147–172.
- 21 Kistler, M. (2017). Higher-level, downward and specific causation. In: M. P. Paoletti, & F.  
22 Orilia (Eds.), *Philosophical and scientific perspectives on downward causation*.  
23 Routledge.
- 24 Kroedel, T, & Schulz, M. (2016). Grounding mental causation. *Synthese*, 193(6), 1909–  
25 1923.

- 1 Leuridan, B. (2012). Three problems for the mutual manipulability account of constitutive  
2 relevance in mechanisms. *British Journal for the Philosophy of Science*, 63(2), 399–427.
- 3 Leuridan, B. (2017). Supervenience: Its logic and its role in classical genetics. *Logique et*  
4 *Analyse*, 198, 147–171.
- 5 Leuridan, B., & Lodewyckx, T. (2020). Diachronic causal constitutive relations. *Synthese*,  
6 198, 9035–9065.
- 7 Levine, J. (1983). Materialism and qualia: The explanatory gap. *Pacific Philosophical*  
8 *Quarterly*, 64, 354–361.
- 9 Lewis, D. (1986). *Philosophical papers: Volume II*. Oxford University Press.
- 10 List, C., & Menzies, P. (2009). Non-reductive physicalism and the limits of the exclusion  
11 principle. *Journal of Philosophy*, 106(9), 475–502.
- 12 McLaughlin, B., & Bennett, K. (2021). Supervenience. In E. N. Zalta (Ed.), *The Stanford*  
13 *Encyclopedia of Philosophy*. Stanford: Metaphysics Research Lab. Retrieved from  
14 <https://plato.stanford.edu/archives/sum2021/entries/supervenience/>
- 15 Mackie, J. L. (1974). *The cement of the universe*. Oxford: Clarendon Press.
- 16 Pearl, J. (2000). *Causality*. Cambridge University Press.
- 17 Schmaltz, T. (2008). *Descartes on causation*. Oxford University Press.
- 18 Spirtes, P., Glymour, C., and Scheines, R. (1993). *Causation, prediction, and search*.  
19 Dordrecht: Springer.
- 20 Stein, N. (2012). Causal necessity in Aristotle. *British Journal for the History of Philosophy*,  
21 20(5), 855–879.
- 22 Stern, R., and Eva, B. (2023). Anti-reductionist interventionism. *British Journal for the*  
23 *Philosophy of Science*, 74(1), 241–267.
- 24 Suppes, P. (1970). *A probabilistic theory of causality*. Amsterdam: North-Holland.
- 25 Woodward, J. F. (2003). *Making things happen*. Oxford: Oxford University Press.

- 1 Woodward, J. F. (2008). Response to Strevens. *Philosophy and Phenomenological*  
2 *Research*, 77(1), 193–212.
- 3 Woodward, J. F. (2015). Interventionism and causal exclusion. *Philosophy and*  
4 *Phenomenological Research*, 91(2), 303–347.
- 5 Woodward, J. F. (2016). The problem of variable choice. *Synthese*, 193(4), 1047–1072.

DRAFT