Mental Causation for Mind-Body Dualists

Antonella Corradini†
antonella.corradini@unicatt.it

ABSTRACT

Interactive dualism is notorious for supporting genuine and autonomous mental causation that is allegedly impossible for its confliction with basic principles of physics. The purpose of this essay is to show the invalidity of this commonplace view, by arguing to the contrary in three different steps. First, I will deal with the objection about the non-scientific character of interactive dualism, as it is conceived of in present-day philosophy of mind. Second, I will illustrate and critically examine three contemporary models of dualistically understood mental causation. Finally, I will summarise the chief assumptions made in these models and try to lay down my own proposal.

Keywords: mental causation, causal closure, interactive dualism, emergent dualism, counterfactual theory of causation, powers theory of causation, difference-making cause

1. Introduction

Mental causation is quite a complicated matter, and this holds true not only for its dualistic variant.[1] However, this variant has the additional task of showing how genuine and autonomous mental causation is possible, and why mental causation, thus understood, does not conflict with fundamental scientific principles. As the objection about the non-scientific character of interactive dualism has already been raised against Descartes, this essay will start by considering this classical theme, as it is conceived in present-day philosophy of mind. Then, I will illustrate and critically examine three contemporary models of dualistic mental causation. Finally, I will summarise the chief assumptions made in these models and try to lay down my own proposal.

[1] For an overview of recent debates on this topic see Gibb (2014).

2. Is Interactive Dualism Incompatible with Fundamental Principles of Physics?

It has often been put forward that mind-body dualism would violate the laws of conservation of energy. The argument against dualism develops as follows. Dualism violates the principle of causal closure.[2] The principle of causal closure is implied by the laws of conservation of energy. Thus, dualism is incompatible with the principles of conservation of energy. Now, since this essay regards mental causation within the framework of a dualistic view, my primary interest is to clarify the meaning, and to establish the correctness, of the argument about causal closure that rests upon the principles of conservation of energy. In fact, every form of dualistically understood mental causation should be conceived as a kind of influence that mental events exert upon the world. But, if this implies a violation of the laws of conservation of energy, basic principles of physics would cast doubt on the very possibility of mental causation. The preliminary task of this essay is, therefore, to state if mental causation is actually in contrast with the laws of conservation.

This issue is addressed by Papineau 2000 and 2002. Both of these works have been very clearly critically analysed by Gibb (2010, and in other places). Here, I follow her analysis. The argument put forward by Papineau comes from his in-depth examination of this issue throughout the history of modern science. Indeed, he begins his investigation by posing the fundamental question: Is it necessarily true that, if the laws of physics are conservative, then the states of physical systems are completely determined by the states of prior systems?

In Papineau 2000 (p. 185), he affirms:

> My original thought was that the completeness of physics would follow from the fact that physics can be formulated in terms of conservation laws. If the laws of mechanics tell us that important physical quantities are conserved whatever happens, then does it not follow that the later states of physical systems are always fully determined by their earlier physical states? Not necessarily. It depends on what conservation laws you are committed to.

[2] In the next sections the principle of causal closure will be formulated in a formal way, and its meaning will be thoroughly examined. In its usual formulation, it reads that all physical effects have only physical sufficient causes. When the principle of closure has this meaning, it coincides with the principle of the completeness of physics, according to which all explicable physical events are such through the laws of physics.

The following pages of Papineau's article consist of a detailed analysis of the development of physics since Descartes' time, serving to support a negative answer to the fundamental question. Through this analysis, he highlights the fact that the conservation laws, by themselves, do not rule out the possibility of non-physical causes. These are ruled out only under the condition that mental forces do not exist. If these existed, they could influence our behaviour, without violating conservation laws, because they, themselves, like other forces, could be conservative. Nevertheless, in spite of the insufficiency of the conservation laws alone to guarantee the physical closure of the world, Papineau argues in favour of the closure, relying on the fact that modern science excludes the existence of mental forces.[3] Ultimately, therefore, after having established the relation between conservation, closure, and existence of non-physical forces, Papineau comes to the final conclusion that, in the present state of science, causal closure follows from conservation. This argument is summarised as follows by Gibb (2010, p. 370):

1. (*Conservation*) Every physical system is conservative or is part of a larger system that is conservative (where a system is conservative if its total amount of energy and linear momentum can be redistributed, but not altered in amount, by changes that happen within).[4]
2. (*Energy*) There is no non-physical energy.
3. (*Exclusion*) No physical effect has a non-physical cause.

Now, Gibb says, the argument – as it is – is not conclusive. This is true, quite apart from the mere fact that there are no mental forces. Indeed, the conclusion (*Exclusion*) cannot be obtained from the premises alone, (*Conservation*) and (*Energy*). It needs two further premises that Gibb defines in the following way:

[3] Papineau offers two arguments in favour of the exclusion of mental forces: «1. The argument from fundamental forces. The first argument is that all apparently special forces characteristically reduce to a small stock of basic physical forces which conserve energy. Causes of macroscopic accelerations standardly turn out to be composed out of a few fundamental physical forces which operate throughout nature. So, while we ordinarily attribute certain physical effects to "muscular forces", say, or indeed to "mental causes", we should recognize that these causes, like all causes of physical effects, are ultimately composed of the few basic physical forces. 2. The argument from physiology. The second argument is simply that there is no direct evidence for vital or mental forces. Physiological research reveals no phenomena in living bodies that manifest such forces. All organic processes in living bodies seem to be fully accounted for by normal physical forces" (Papineau, 2000, pp. 197–198).

[4] For this formulation of the principle of conservation see Gibb (2010, p. 367), where the author also mentions Broad (1925, p. 105).

Physical Affectability: The only way that something non-physical could affect a physical system is by *affecting* the amount of energy or momentum within it, or by *redistributing* the energy and momentum within it.

Redistribution: Redistribution of energy and momentum cannot be brought about without supplying energy or momentum.

Let us note that, in the event that mental forces existed, they could act precisely in one of these two ways (either affecting the amount of energy or redistributing the energy) and yet, being conservative, avoid violating the conservation laws. But, Gibb asks herself, why should these two principles be *necessary* to affect physical events?

Actually, both principles may be justified if we consider the relation of causation as energy transfer. From this concept, in fact, both the principle of *Physical Affectability* and that of *Redistribution* would immediately follow. However, introducing the theory of causation as energy transfer in order to justify the principles at stake is an implausible move for two reasons. First, it is not reasonable to assume the theory of energy transfer, if the purpose behind it is to show that the causal closure follows from the principles of *Physical Affectability* and *Redistribution*; in fact, closure follows on directly from the theory of energy transfer. Indeed, if the act of causation between a cause and an effect consists of energy transfer, and energy is only physical energy, the cause of a physical event cannot be other than a physical event, so the physical world is causally closed. Second, there are strong reasons against the identification of the causal relation as the transfer of physical energy. Even though the phenomena of physical causation occur by means of physical energy transfer, this transfer is the way in which the physical relation manifests itself, not the relation of causation itself. The latter – as we will see better in the second section of this essay – is either a relation of regular succession, or of production among events, which does not necessarily imply that energy transfer also occurs. It may be that energy transfer occurs, and that it is nomologically necessary, because the very nomological regularity that is expressed by the causal relation involves energy transfer, or its transformation from one form to another.

It is noteworthy that just undermining Papineau's argument is sufficient to call into question the principle of *Redistribution*. Let us simply assume that the mental cause could lead to a redistribution of energy in the physical world, without having to provide in turn an amount of new energy. The denial of the

principle of *Redistribution* is explicitly made by Broad when he claims that:

> ...all the energy of our bodily actions comes out of and goes back into the physical world, and [...] minds neither add energy to nor abstract it from the latter. What they do [...] is to determine that at a given moment so much energy shall change from the chemical form to the form of bodily movement; and they determine this, so far as we can see, without altering the total amount of energy in the physical world (1925, p 109).

In conclusion, Papineau's argument is defeated because the principles of *Physical Affectability* and of *Redistribution* are both fraught with difficulties. Of course, having shown that the argument is inconclusive neither means having shown that the phenomenon of mental causation is a fact, nor that it may be considered to be free from difficulties. I will deal with this topic in the next two parts of this essay.

3. Models of Dualistic Mental Causation

3.1 Uwe Meixner's Interactive Parallelism

I begin my exposure of currently held models of interactive dualism with Uwe Meixner's proposal, because he explicitly deals with the conceptions of causality underlying the mental causation debate.[5] Having even a sketchy knowledge of general theories of causation is quite important, since these allow us to comprehend better the true nature of the debate on mental causation, as the dualists conceive it.

3.1.1 Meixner's Theory of Mental Causation

In what follows, I will mention the main topics of Meixner's theory of mental causation, such as the nature of the causal *relata*, the character of the causation relation, the analysis of causal overdetermination, the non-transitivity of the causation relation, his criticism of Lewis's counterfactual theory, and his evaluation of the probabilistic theory of causation.

[5] On this, see Meixner (2001).

3.1.1.1 The Nature of the Causal-Relata

Meixner maintains that there is no significant distinction between event-causation and state-of-affairs-causation. At most, he affirms that event-causation can be deemed a special form of state-of-affairs-causation (Meixner, 2004b, p. 352). Also, the causal *relata* can be conceived of as Kimean events – that is, as the instantiation of a property by a substance at a time. However, like Lowe (2013), he also accepts kinds of agent-causation besides event-causation (2004b, pp. 352 ff). In such cases, the causes are agents, while the effects are still events. In what follows, this distinction is not relevant, for an agent's causal action can be interpreted as the influence that an agent exerts through the activation of one of her powers. Moreover, the states-of-affairs/events that constitute the causal *relata* are formalised through variables x, y, z...

3.1.1.2 The Nature of the Causation Relation and Causal Overdetermination

Meixner supports a theory of causation that is implicitly based upon nomological regularities. He defines the causation relation as follows: «x causes y [...] $=_{def}$ x and y are actual events, x has a temporal antecedent and occurs before y occurs, and in every nomologically possible world in which x occurs, y occurs, too [...]». The latter condition is logically equivalent to the condition «the totality of the laws of nature has the logical consequence that the occurrence of x implies the occurrence of y» (original English version of Meixner 2006, pp. 1779–1780).

Clearly, this is a necessitarian theory of the causation relation xCy, in which x performs the function of the sufficient condition of effect y. That causation is thus understood is important for various reasons. First, this is a feature not just of Meixner, but also of Lowe's dualistic model of psycho-physical causation (Lowe, 2008).[6] I believe that this similarity is not coincidental. In fact, a theory in which causation is a sufficient condition guarantees, on the one hand, the efficacy of causation, but, on the other, it does not exclude the action of a different sufficient cause. This is seen in the case of mental causation, in which the occurrence of an instance of mental causation is usually accompanied by the occurrence of a physical instance. Of course, the co-occurrence of two sufficient causes raises the suspicion that one of them is redundant – that

[6] In her most recent essay on this topic, Gibb distances herself from the theory of the cause as sufficient condition. For this, see Gibb (2015).

is, that causal overdetermination occurs. This problem would not arise for the supporters of any theory of causation as *conditio sine-qua-non* – for instance, a counterfactual theory of causation. On this theory's construal, different causes could contribute, all in a relevant way, to the determination of the same effect. The co-occurrence of both causes would not be a case of overdetermination. Overdetermination takes place when both causes are sufficient.

However, Meixner considers that to be a normal aspect of the causation relation. After all, given his nomological theory of causation, from xCy obtaining iff all x worlds are also y worlds, it naturally follows that all $x \wedge z$ worlds are also y worlds and, thus, that $x \wedge z$ is a cause of y in the same way as x is a cause of y. Meixner (2006, p. 1780) asks: «Is this bad? It can seem very bad if there is a great incongruence between z and y regarding content and significance – but only if one forgets that y is anyhow nomologically necessary. There seems nothing particularly unacceptable in every actual event z before y causing y if y is nomologically necessary anyhow». The only kind of causal overdetermination that is reasonably deniable occurs when two causes – both sufficient – are reciprocally independent, meaning that they do not have a common cause, and do not cause each other. Perhaps this kind of overdetermination is false, but its putative falsity cannot be found out a priori. As Meixner notes: «A priori reasons that might be adduced against causal overdetermination have more to do with us – with the intellectual economy of the production of causal explanations – than with objective reality [...]» (p. 1781).[7]

3.1.1.3 Non-Transitivity of the Causation Relation

In (2004b, pp. 365–366), Meixner maintains that event-causation is not provably transitive and that the paradigm of truly efficient causation[8] that has guided the analysis of event-causation in his paper is indifferent on this issue. Agent-causation – he argues – is trivially intransitive, owing to the categorical difference of cause and effect. Thus, «from the point of view of truly efficient causation, event-causation – its analog – may, therefore, as well be not transitive». This analysis is useful for emphasising the difference – as Meixner himself does – between his own viewpoint and Lewis's. In a footnote, he writes:

[7] For clarity sake, in what follows I will name the first kind of overdetermination "weak overdetermination" and the second "strong overdetermination".

[8] On Meixner's theory, agent-causation is a paradigm for causation in general. The causation that is modelled according to agent-causation is named "Truly efficient causation".

> Lewis insists that "causation must always be transitive" [...], whereas he allows symmetrical and reflexive instances of event-causation (Lewis, 1986, pp. 167 and 213). Seen from the point of view of the paradigm of truly efficient causation, however, event-causation can very well be not transitive, whereas its not being asymmetrical or not irreflexive is as absurd as can be. [...] we have: nothing can make itself actual, nothing that makes something actual is made actual by that which it makes actual. And event-causation, being the analog of truly efficient causation, has better yield the analogs of these principles.

3.1.1.4 Criticism of Lewis's Counterfactual Theory

Meixner rebuts Lewis's *sine-qua-non* counterfactual theory of causality. The reason for this rebuttal is the similarity relation, which is very obscure and exposed to the uncertainties of imagination. In fact, Meixner poses the following question:

> But which criteria of comparative similarity between possible worlds should we follow? There is no objective answer to this question: everybody can choose the criteria of comparative similarity he or she likes. For example: Is a possible world more similar (than another) to the actual world if it is more similar to the actual world with regard to phenomenal appearance, or is a possible world more similar to the actual world if it is more similar to it with regard to nomological structure? [...] The counterfactual sine-qua-non concept of causation is not an objective concept, but depends (regarding its content) to a very high degree upon the beliefs and volitions of human subjects and is, therefore, in this sense, a subjective concept (original English version of Meixner 2006, p. 1777).

3.1.1.5 A Comparison between Meixner's Nomological View and the Probabilistic Theory of Causation

A comparison between Meixner's nomological view and the probabilistic theory of causation can help to highlight both the difference between the two models, and the extensibility of the nomological theory to its probabilistic variant. This comparison will be useful later, when we will have to deal with Lowe's model, which uses the concept of probability to illustrate the impact of mental causes on physical ones. As is known, the way in which the probabilistic theory of causation sees a cause as contributing to the effect, is by raising the probability of

the latter. Clearly, the probabilistic theory conceives of the causal relationship non-deterministically; thus, it features in a different framework from the deterministic, which is typical of the nomological theory. In the nomological theory, there is a total determination between the cause and its effect, both because the cause is a sufficient condition, and because the event-effect follows in all the worlds in which the event-cause occurs. By contrast, a probabilistic implication between two events exists when determination is partial. The event-effect only occurs in some of the worlds in which the event-cause occurs. Does this mean that the cause is understood as a sufficient condition of the effect? Certainly not. However, the cause can be understood as that which contributes to produce the effect, if the partial determination is conceived of in propensionalistic terms.
This is the perspective Meixner endorses (2001, p. 272). It is grounded on the concept of an object's x propension at t to produce the effect of being G at t'. Such a propension is named by Meixner the "essential propension" of x in t. This can be increased by the fact that x is to possess in t a further property F. In this case, F is a probabilistic cause in t. Is it a sufficient cause? Certainly not, by itself, but, combined with x's essential propensity to produce G, it becomes a sufficient condition, either because the final propension reaches 1, or because the effect is random, in which case it is caused, though not deterministically, by x's propensity to produce G as essentially x and as F. Clearly, underlying the propensional perspective of causation, is the thought that causation is a contribution to production – that is, reduction of resistance to the realisation (2001, p. 272). This does not apply to the regularity theory that is the basis of the nomological theory of causation. Yet, both theories are compatible, since the nomological is a borderline case of the theory of partial determination. The latter, in its propensional interpretation, provides the possibility of extending the theory to encompass the case of production in random conditions. The step of reconciling the notion of causation with that of libertarian action is easy.

3.1.2 Meixner's Parallelistic Model

3.1.2.1 *Causal Closure Principle*

We can now proceed to illustrate Meixner's model. It is an interactive model, in which mental causes exert a real influence on human action. For this reason the author has to face the problem of the physical world's causal closure. Does mental causation violate the causal closure of the world? And, if so, which principle,

among many, is violated? And how is such a violation possible? Meixner devotes much space to the analysis of the causal principles (2004a, chapter 8). First, he believes that all of these principles have a metaphysical nature and, for this reason, that they should be justified or refused on the background of metaphysical considerations. Therefore he refuses the following strong principle of closure:

1. *Strong closure principle PCC2* (also named *Principle of exclusion of mental causes*):[9] No physical effect has a non-physical cause.

$$\forall x \forall y (Fy \wedge xCy \rightarrow Fx)$$

This principle is not grounded on the physical laws of conservation of energy and momentum, but is a metaphysical principle characterising physicalism as a philosophical doctrine (2004a, pp. 301–305). By contrast, Meixner believes that the following weak principle of closure is metaphysically justified:

2. *Weak closure principle PCC1* (also named *Principle of completeness*): Every physical effect has a sufficient physical cause.

$$\forall y [Fy \wedge \exists z (zCy) \rightarrow \exists x (Fx \wedge xCy)]$$

At first sight, it may seem that even *PCC1* is incompatible with interactive dualism. In fact, the principle reads that, if a physical event – such as an arm's movement – has a cause, then it has a physical cause. But, since the variables that stand for the cause-event are existentially bound, the principle does not entitle us to say that the cause must coincide with the physical cause. There may be both a mental cause and, at the same time, a physical cause. Even if these were two *sufficient* causes, one of them – the mental – might be purely epiphenomenal. Meixner, however, avoids the difficulties arising from this issue in virtue of his theory of the physical causal representation of a mental event (2004a, p. 311).

3.1.2.2 Theory of the Physical Causal Representation of a Mental Event

This theory envisages a kind of parallelism between mental events and their neural correlates. This parallel psycho-physical structure can be illustrated as

[9] It is the same principle mentioned in Section 1.

follows: let y be a variable for mental events, and x a variable for physical events and either $z=y$ or $z=x$. Thus, the relation $x\mathrm{R}^*y$ symbolises that x is a causal representative of y. Its definition is the following:

$xR^*y =_{def} \forall z(zCx \leftrightarrow zCy) \wedge \forall z(xCz \leftrightarrow yCz)$
$=_{def}$ the causes of x are also causes of y and the effects of x are also effects of y

Clearly, $x\mathrm{R}^*y$ is a symmetrical relation. The figure below shows the one-to-one parallelism between the set of the mental events and the set of the correlated physical events. In the scheme, the double vertical arrow expresses the relation of correspondence, while the oblique lines indicate the relation of causation (x_1Cy_2, x_1Cx_2, y_1Cy_2, y_1Cx_2...). It should be emphasised that on Meixner's reading the one-to-one relation of correspondence is not a causal relation. A neural firing does not cause a mental event – nor does the opposite occur – even though a relation of reciprocal representation does exist between the two of them. Note that, while no causal relation exists between the x_i and the y_i, this relation does exist between x_i and all the x_{i+1} and the y_{i+1}, as well as between the y_i and all the x_{i+1} and the y_{i+1}.

y_1		y_2		y_3		y_4		y_5	Mental
⇕	╳	⇕	╳	⇕	╳	⇕	╳	⇕	
x_1		x_2		x_3		x_4		x_5	Physical

The correspondence between conscious mental events and neural events is expressed by the following *C4* principle (Meixner, 2004a, p. 297):

> *C4*:For every conscious event y there is a physical event x such that x is a causal representative of y.

The principle can be formalised as follows:

$$\forall y(My \rightarrow \exists x(Fx \wedge yR^*x))$$

On the basis of C4, the causal closure principle PCC1 comes to have the following meaning (p. 303):

$$\forall y[Fy \wedge \exists z(Mz \wedge zCy) \rightarrow \exists x(Fx \wedge xCy)]$$

Now, the theory of causal representation allows us more deeply to understand the meaning of *PCCI*. For Meixner, the relation of causation indicated by letter *C* is endowed with a temporal index, which is the same for the two occurrences of *C* in the principle. For this reason the true meaning of the principle is made explicit by the following formula:

$$\forall t \forall y[Fy \wedge \exists z(zC_t y) \rightarrow \exists x(Fx \wedge xC_t y)]$$

The principle reads that if there is a cause of y at t, then there is also a physical cause of y at t. Given the relation of causal representation between physical and mental events, the principle is clearly justified.

3.1.3 Critical Remarks

The theory of causal representation considers weak causal overdetermination to be a normal phenomenon. Indeed, physical causes and mental causes are both sufficient, even though they depend on each other. This is a problematic aspect of Meixner's conception. This issue will be further developed in the third section of the present essay. A second critical aspect of Meixner's theory is his nomological theory of causation. I do not argue that this theory is wrong *per se*. The arguments made by the author are well founded and comprehensive, but this does not remove the possibility that this theory of causation could, or should, be supplemented by a theory of causation as production. This integration would allow the nomological theory to respond to instances of agent-causation – of which Meixner himself is an advocate. Meixner's theory of sufficient causation is, in fact, a necessitarian theory based on the concept of natural law, and not a theory of causation as production. But it is a phenomenological datum of primary importance that, when we – as agents – feel responsible for being the cause of a physical effect in the world, we experience being the producers of that effect. A theory of mental causation should take this into account, and not be limited to detecting a regularity – however necessary – between mental events and physical events.

3.2 E. Jonathan Lowe's Emergent Dualism

While Meixner intends to show that the principle of causal closure is no threat to interactionism in virtue of its essentially metaphysical nature, and of the compatibility of its weak formulation with his parallelistic model, Lowe is committed to showing that there is no appropriate formulation of the principle that – at the same time – is both sustainable and incompatible with his emergentist model. He analyses a wide range of principles of differing strengths, with the intention of showing that, on the one hand, they are not too weak to confute the argument of causal closure nor, on the other, too strong to be rationally adopted through scientific or metaphysical arguments (Lowe, 2008, p. 43). The most important among these principles are *IG* and *IH*.

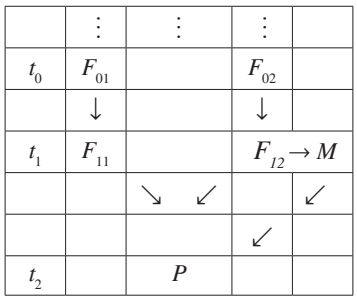
3.2.1 The Principle of Closure IG

Principle *IG*: At every time in which a physical event has a cause, it has a sufficient physical cause.

The principle is based on three ideas: (i) The cause is conceived of as sufficient cause; (ii) a cause which is *prior* to the mental cause can also be considered a sufficient cause of a physical event; iii) the prior physical cause of the event can be placed at a different time from that in which the mental cause acts. The principle of closure *IG* has, then, the same shape as that of Meixner's principle *PCC2*:

$$\forall y[Fy \wedge \exists z(zCy) \rightarrow \exists x(Fx \wedge xCy)]$$

However, the principle *IG* is satisfied by situations of a different type from those covered by Meixner's parallelistic model. Let us consider the following example. Let us imagine that *P* stands for a physical event, such as the rising of an arm. Let us suppose that the rising of the arm can be explained through the intention that the person who lifted the arm had, to attract someone's attention. Clearly, this intention is the mental cause of the arm's rising. Yet, the arm's rising is accompanied by a series of chains of neural events that makes the arm's rising possible. The whole psycho-neural process corresponding to the intentional act can be illustrated in the following way:



What information can we draw from this scheme? It states that P at t_2 is the effect of two sufficient causes acting, albeit at different times. The temporally immediate cause is constituted by the set of conditions acting at time t_1. This is clearly a – at least partial – mental cause. At time t_0, however, what is operating is the totally physical cause constituted by the set of conditions $\{F_{01},...,F_{02}\}$. If overdetermination occurs here, it is of a weak kind, since the causes operating at different times are reciprocally dependent. However, we need to ask what the role of M is, among the other components of the cause at t_1, and whether there is overdetermination between M and the other components of the cause. Lowe reflects upon this carefully, trying to outline the difference between M and the physical causes (2008, pp. 47 ff). In the light of his emergent dualism, M's emergence from a set of physical conditions does not yield any difficulty. The problematic aspect would, rather, be M's affecting the causal process, without a physical support. This would occur if M's causal action were not simultaneously accompanied and supported by a physical correlate. This is the reason why Lowe locates M at the same time in which its correlate F_{12} occurs. Nevertheless, the way in which F_{12} is related to M remains a critical point of Lowe's model. Is M caused by F_{12}? It seems so, because Lowe argues that F_{12} causes P partially through causing M. But why must F_{12} pass through M in order to cause P? An answer to this question is at the heart of Lowe's model.

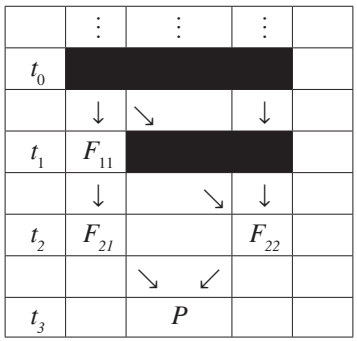
3.2.2 What Are Mental Causes For?

Lowe's view that mental causes are not epiphenomenal, although they are dependant on the physical, is strengthened by his further claim that mental causes do not have the function to explain why certain events occur, but rather to explain why the occurrence of these events is not coincidental. Also, this claim clears up why the explanatory function of mental facts is invisible to science.

With regards to the example of the arm's movement, Lowe remarks that the bundle of causal chains behind this voluntary act presents a high degree of complexity: this depends on its tree-structure that, in moving from an interwoven and diffuse ramification, converges towards the unity of the arm's movement. Now, from the physical point of view, this convergence appears wholly coincidental. As Lowe puts it: «[...] as physical science traces back the physical causes of our bodily movements into the maze of antecedent neural events, it seems to lose sight of any unifying factor explaining why those apparently independent causal chains of neural events should have converged upon the bodily movements in question» (Lowe, 2000, p. 581). On Lowe's construal, a mental cause has the function of explaining the fact that this convergence occurs; this explanation, on the other hand, is made possible by the intentional nature of a mental event. In Lowe (2008, pp. 28–29) the same concept is illustrated through the example of a man's death owing to the fall of a slate from the roof of a house. Event *P*, constituted by person *B*'s violent death, can be described as the point of convergence of two fundamental causal chains, F_{-2} and F_{-1}. They represent, respectively, the causal chains leading to the slate's fall and those determining *B*'s stroll up to the house from which the slate falls.
Description of event *P*:

	⋮	⋮	⋮	
t_0	F_{01}		F_{02}	
	↓		↓	
t_1	F_{11}		F_{12}	
	↓		↓	
t_2	F_{21}	slate's fall	F_{22}	
		↘ ↙		
t_3		*P* (death of *B*)		

In the given description, *B*'s death appears to be coincidental. *P* is the convergence point of two independent causal chains, leading, respectively, to the slate's fall (F_{-2}) and to *B*'s stroll up to the house from which the slate falls (F_{-1}). This convergence is coincidental, unless a further cause occurs, which is the intention of *Z* – an enemy of *B* – to provoke his death. The inclusion of the intentional cause represented by *M* explains the occurrence of the event in a way that takes away its coincidental character. See the Figure below:



This second figure suggests that M represents the murderous intent of Z. After detecting the beginning of B's walk at time t_0, Z acts at time t_1, to make sure that, starting from the position of the slate expressed by F_{12}, the slate falls from the roof at t_2 and hits B at time t_3 at which B passes by. Let us ask, first of all, whether this way of understanding the function of mental events is at least compatible with *IG*. Can we say that at t_1 the cause of F_{22} is constituted by the chain of the prior causes? Obviously not. At t_1 the sufficient cause of F_{22} must also include M. But M is a mental cause, whose physical cause occurs at a prior time. Thus, the closure principle *IG* is satisfied: assuming that the causation relation is transitive (meaning that causes of the same effect can operate at different times), we have:

$$\{F_{12}, M\} C_{t_1} F_{22} \wedge \{F_{01}, F_{02}\} C_{t_0} F_{22}$$

This example is also useful in accounting for the way in which mental causes are invisible to science. Lowe argues that physical science traces the physical causes of our bodily movements back to the maze of antecedent neural events. Therefore, physical science loses sight of any unifying factor that explains why those causal chains of apparently independent neural events have a convergence point (Lowe, 2008, p. 53).

3.2.3 Critical Remarks

As we have just seen, the specific role attributed to mental causes by Lowe's model manifests his refusal to conceive of mental causes as being reducible to physical causes. But I would like to ask whether his proposal is entirely satisfactory, and to this end I will pose three orders of questions. First, can we take for

granted that the causation relation involved in the example of the intentional killing is transitive, according to its definition? That this relation is transitive in the example of the arm's movement dealt with previously is quite obvious, because it is a deterministic relation. A deterministic cause of the arm's movement exists both at t_0 (the physical cause) and at t_1 (the partially mental cause). By contrast, the example of the intentional killing is placed into a probabilistic framework. In fact, *M* has been introduced in order to avoid that the slate's fall and the subsequent death of *B* is coincidental, that is, an event characterised by low probability given the initial conditions. But then the causation relation in virtue of which $\{F_{01},F_{02}\}$ is a sufficient cause of F_{22} at t_0 is not the same in virtue of which $\{F_{12},M\}$ is a sufficient cause of F_{22} at t_1. The first is characterised by a probability nomologically less than 1, the second by a probability equal to 1. Thus, transitivity does not hold true.[10]

A second problematic aspect arises from the fact that, according to Lowe, the unifying role of mental causes is compatible with an even stronger formulation of the closure principle. Analysing this formulation briefly is worthwhile, because it allows us to consider the difference between *event causation* and *fact causation*, which is a central claim of Lowe's theory. Lowe first introduces the concept of transitive causal closure (2008, p. 54): the transitive causal closure of an event means the totality of the causes on which that event depends.[11] The principle of closure can then assume the following form:

> *IH*: The causal closure of every physical event contains only other physical events in its transitive causal closure.

So, what is the sense of the claim that the unifying function of mental causes – and the ensuing dualistic interaction – is compatible with *IH*? Of course, this principle rules out that a mental event *M* can be part of the transitive causal closure of a physical event – it is neither a cause among physical causes $F_1 \ldots F_n$ nor can it be conceived of as the cause which triggers the succession of causes $F_1 \ldots F_n$. However, *IH* does not rule out that *M* explains the *fact* that $F_1 \ldots F_n$ converge into the causation of *P*. In other words, *M* is responsible for the *fact* that *events* $F_1 \ldots F_n$ cause *event P*. It is at this point that Lowe makes an audacious analogy between

[10] Meixner's model does not present this problematic aspect, since either event *z* occurs at the same time as event x, or because event z is the same event as x, or because it is its physical representative.

[11] It is not important for us to put or not to put the additional condition that as a whole these causes are also sufficient.

mental and divine causation. Just as one can conceive of God as being the cause of the entire infinite series of mundane causes – because He explains the occurrence of the series – without being a cause among the others (not even the first), one can also conceive of the mental cause as being the cause explaining the fact that the entire causal closure of event *P* occurs (Lowe, 2008, p. 55).

Although this analogy is enlightening, it highlights other problematic aspects of Lowe's model. First, to which series of physical causes should the series of causally linked temporal events created by God correspond? Perhaps to those causes preceding event *P* since the beginning of the universe? But how could a mental cause be the grounding of this series? Only a strongly idealistic view of mental events could justify this claim. Second, the mental cause is – unlike the divine – an immanent cause, which means that, if the mental cause is efficacious, it is so together with other physical causes, thus it should be part of the entire transitive causal closure, in violation of *IH*.

Finally, a third critical aspect of Lowe's model regards the author's emergentist view of mental causation. As already mentioned, he commits himself to a kind of emergent dualism. It is a version of dualism, «which preserves one central tenet of physicalism, namely, that every physical event has a set of wholly physical causes which are collectively causally sufficient to the occurrence of that event» (2008, p. 40). Since Gibb supports a cognate kind of emergent dualism, I will discuss both authors' views in the final part of Section 3.

3.3 Sophie C. Gibb's Double Prevention Model

Gibb's (2015) article on the principles of causal closure lays down the general theoretical background against which her model of mental causation is placed. This background is marked by the fact that she distances herself from the notion of mental cause as sufficient cause. She refuses «the assumption that every physical event that has a cause has a sufficient cause, or, at least, that every physical event that has a cause has a cause that is sufficient to fix its chances» (Gibb, 2015, p. 4). In Gibb's opinion, there are good reasons to think that not all of the physical effects have a set of causes that collectively constitute a sufficient condition for their existence. This claim does not derive from reflections on quantum physics – as the reasons put forward by Gibb for refusing the idea of sufficient causation do not only concern forms of deterministic causation – but refer to kinds of causation in which the cause fixes the probability of the effect. Rather, Gibb's reasons are grounded in the powers theory of causation

that the author herself adopts. She argues that:

> The crucial point is that, given this account – and, indeed, given any version of the powers theory of causation – the existence of all of the contributory causes of an event (that is, the existence of its complete cause) is not always sufficient for the existence of that event. The central examples that demonstrate this point involve cases of double prevention (Gibb, 2015, p. 7).

Double prevention occurs when an event that would prevent another event from having a certain effect is itself prevented from doing so. Now, the phenomenon of double prevention can be explained through the powers theory of causation. To this end, Gibb introduces the following example of double prevention. A barrier is placed in front of a vase to protect it from breaking. However, the barrier is wired up to an explosive device, which will blow up the barrier if a button on the device is pressed. Therefore, this button plays the role of the double preventer, as it prevents the action of the barrier that has the function of preventing the breaking of the vase. In dispositional terms, the same concept can be illustrated as follows. Disposition *A* may be disposed to prevent the manifestation of disposition *B*, either because *A*'s manifestation results in the loss of *B,* or because it merely blocks *B*'s manifestation. In the previous example, by pressing the button and so eliminating the barrier, one prevents the manifestation of its protective power, which would prevent the breaking of the vase.

Gibb notes that

> what is of crucial significance for the purpose of this paper is that the device's button being pressed cannot be a contributory cause of the breaking of the vase according to the powers theory of causation. More generally, given the powers theory of causation, a double preventer event *cannot be a cause* of the event that it has prevented from being prevented (2015, p. 9).[12]

What is the reason for this?

> In brief, this is because *absences* cannot be causes according to the powers theory of causation, for an absence cannot bear powers and hence cannot be disposed to act in any way. Given that absences are not causes, there cannot be *a chain of unbroken causation* from the double preventer event to the event that it has prevented from being prevented. Hence, in our example,

[12] For this also see Mumford and Anyum (2009).

> the pressing of the button causes the destruction of the barrier, but given the powers theory of causation, the barrier's destruction cannot in turn be a cause of the vase's breaking, for this is really just to say that the *absence* of the barrier is a cause of the vase's breaking. Therefore, given the powers theory of causation, the pressing of the button is *not a cause* of the vase's breaking, for there is not a chain of unbroken causation from the pressing of the button to the breaking of the vase. More generally, given the powers theory of causation, double prevention is not causation (2015, p. 9).

Now, according to the powers theory of causation, a mental cause of P is a double preventer of P. On Gibb's reading, that the mental cause is a double preventer of the physical effect is relevant to its causation. Nevertheless, a double preventer is not a true cause, since an absence does not have the power to produce anything.

In order to illustrate Gibb's view on this point, I take the example of the voluntary arm's movement she herself analyses and comments on (2015, p. 15). Let us imagine that Fred has the desire to raise his arm. Let us denote his desire with m_1 and the event consisting in the firing of neuron 1 – correlated with such a desire – with n_1. b_1 denotes the arm's movement. Also, let us surmise that the occurrence of the arm's movement also requires the firing of neuron 2 – a neural event we denote with n_2. Moreover, n_1 has the power to produce n_2, that is, to let neuron 2 fire, which in turn has the power to move the arm. Therefore, the occurrence of n_1 causes n_2 and n_2 causes b_1. With m_2, instead, we denote Fred's desire to keep his arm at rest, which has the power to prevent n_2 from causing b_1. b_2 denotes the keeping of the arm at rest.

We now have three scenarios.

1. Scenario: the arm's movement.

Let us now hypothesise that Fred's desire to move his arm is stronger than his desire to keep it at rest. What does this imply? It means that the desire m_1 – that is the desire to raise the arm (owing for example to a bad case of pins and needles) – prevents the manifestation of m_2's power to prevent n_2 from occurring, and so to block the arm's movement. In conclusion, m_1 prevents m_2 from blocking n_2 to cause b_1.

The structure of this process can be exemplified by the following figure:

m_1 desire to move the arm for a feeling of pins and needles	$-\circ$	$\boxed{m_2}$ desire to keep the arm at rest	$\cdots>$	$\boxed{b_2}$
		$\vdots$		
n_1 neural correlate of desire	$\rightarrow$	n_2 consequent neuromotor impulse	$\rightarrow$	b_1

where:
m_1 denotes the desire to move the arm;
m_2 denotes the desire to keep the arm at rest;
n_1 denotes the firing of neuron 1;
n_2 denotes the firing of neuron 2;
b_1 denotes the arm's movement;
b_2 denotes the keeping of the arm at rest;
$\cdots\circ$ denotes the unsuccessful act of prevention;
$-\circ$ denotes the successful act of prevention;
$\rightarrow$ denotes causation;
$\cdots>$ denotes blocked causation;
$\boxed{\ldots}$ denotes the prevention of ...

Here, the action of the arm's movement is causally closed, because b_1 is only preceded by physical causes. Moreover, m_1 is not a cause of a physical event, but rather a mental event enabling movement. In fact, it enables n_2 to cause the movement by preventing m_2 from preventing the causation of b_1 on the part of n_2. Actually, even though m_1 contributes to determine b_1, it is not acting on a physical event, but, rather, on another mental event. Clearly, the example shows that the principle of causal closure – in its formulation *IH* – is not violated. Indeed, the causal closure $\{\ldots n_1, n_2\}$ of b_1 only includes physical events.

An apparent objection to this explanatory framework is as follows: how does m_1 succeed in affecting m_2? According to Gibb the reply is very simple: this influence is possible because it takes place among mental events, and the influence of mental events on each other does not imply any exchange of mental energy. Gibb believes, in other words, that the capacity of exerting influence lies not in an exchange of energy, but in the nature of the powers of a mental entity.

A much more serious objection to this interpretation of the arm's movement is represented by the fact that the prevention on the part of m_1 of m_2's action of preventing n_2's action presupposes that m_2 has the power to influence the chain of the physical causes (Gibb, 2015, p. 16). In other words, even if m_1 did not exert its influence, m_2 would still have the power to prevent n_2 from causing b_1. This means that m_2 could affect the chain of physical causes. By deploying the previous figure we can illustrate this case in the following way:

2. Scenario: the arm is kept at rest

m_1 desire to move the arm for a feeling of pins and needles	...○	m_2 desire to keep the arm at rest		
		⏐		
n_1 neural correlate of desire	→	n_2 consequent neuromotor impulse	...>	$\boxed{b_1}$

Gibb rejects this criticism (2013, pp. 202–210; 2015, p. 17) by arguing that this problem could be overcome if, every time in which n_2 and m_2 occur, m_1 is present for preventing m_2 from preventing n_2 from causing b_1. And, that it might have, if there were a neurological event n_0 that implied the existence of n_1 and m_1. On Gibb's view, this reply is wholly coherent with emergent dualism. In fact, this deems mental entities to be dependent on physical entities, and understands such a dependence as a *causal* dependence. The proposed solution thus leads to the following third scenario.

3. Scenario: genesis of the arm's movement

	↗	m_1 desire to move the arm for a feeling of pins and needles	—○	$\boxed{m_2}$ desire to keep the arm at rest	…> absence of causation of b_2 on the part of m_2	$\boxed{b_2}$ keeping the arm at rest
n_0				⋮ ○ prevention of the prevention of the causation of b_1 on the part of n_2		
	↘	n_1 neural correlate of pins and needles	→	n_2 neural firing	→	b_1

In this scenario, the presence of n_0 implies the presence of n_1 and m_1, which does not allow m_2 to fulfil its function of preventing n_2. In this way, the causal chain that could lead to b_2 is excluded, and only the chain leading to b_1 is admitted.

Gibb's reply to this criticism implies the acceptance of a kind of emergent dualism quite similar to Lowe's. Some critical remarks on this, and other aspects of Gibb's model, will be addressed in the next section.

4. Essential Requirements of Dualistic Mental Causation

4.1 Introduction to the Section

In the analysis of the three dualistic models investigated in the previous section, I have taken account of two fundamental orders of consideration. On the one hand, I scrutinised the concept of causality that underlies each of these models and, on the other, the kind of relationship between the domain of mental causes and that of the physical. With regards to the first aspect, I noticed a substantial similarity between Meixner and Lowe's conceptions of causation, as both share a conception of cause as sufficient cause. Gibb departs from this view, because she argues that a collectively sufficient cause of an event (or of its probability if the event is undetermined) may be lacking because the event (or its probability) may depend in an essential way on events of double prevention, and these events are not causes, but simply prevent a causal action. I will dis-

cuss both concepts of causation, and try to defend a theory of the mental cause that takes into account its being a necessary condition, but within a framework that has all the advantages of a theory of the cause as sufficient condition. In this part, I will deploy Menzies' notion of difference-making cause.

As to the second aspect, there is a marked difference between Meixner's model and the two models of Lowe and Gibb. Meixner's model is characterised by a relation of representation between mental events and their neural bases. This model is clear in its formal structure, and is appropriate for the function it has to fulfil. Nevertheless, it deserves to be investigated further, also owing to the fact that the representation relation is the main factor responsible for the phenomenon of overdetermination. The other two models both comply with the emergentist idea that the mental dimension emerges from the physical. Both models, moreover, defend a dualistic version of emergentism – the coherence and plausibility of this will be scrutinised in what follows.

4.2. What Theory of Causation?

A theory of causation capable of supporting an adequate dualistic theory of mental causation should include, in my opinion, both aspects of causation, that is, its being both a sufficient and a necessary condition. John Mackie's theory (1965; 1974) teaches us much in this regard. According to him, a cause is a necessary component of a whole sufficient condition of the effect. The sufficient condition must include all that is necessary at a certain time for the effect to occur. To be sufficient means exactly this. Let us then denote with X all that is necessary, in addition to x, for the causation of y. xCy should then be interpreted as an elliptic expression of the more complex relation $(X \wedge x)Cy$ existing between the sufficient cause $X \wedge x$ and effect y. What is the function of x within $X \wedge x$? It is, as Mackie would say, the function of an INUS condition for the occurrence of y.

This way of analysing the general concept of cause can usefully be applied to the analysis of a mental cause, since it enables us truly to understand the function of a mental event, in the whole complex of the causal chain that is sufficient to create the effect. When a person moves her arm voluntarily, many factors are called upon that contribute to the effect. These factors are of a various nature, from factors of a mechanical kind – which explain the operation of the arm as a lever – to those of a physico-chemical nature – which explain the fact that chemical energy can be transformed in mechanical energy of trac-

tion of the muscles; from neural factors – explaining the transmission of nerve impulses from the centre to the periphery – to factors of a mental kind, which constitute the reasons of an action.

All of these factors, without exception, together contribute to explaining the whole process leading to the final action. The process is broken down into various stages, the sequence of which is explained by the action that the causal factors belonging to a certain phase exert on the next phase, according to a precise temporal scansion. Ideally, it would be possible at every moment to be able to identify a section of the process consisting of all of the sufficient conditions that generate the transition to the next stage. Why sufficient? Because, otherwise, it would not make sense to hypothesise that the final event can be explained. Science requires that events are explicable by means of sufficient causes, even if the causal laws deployed in the explanation are probabilistic laws. In the latter case, the causes are sufficient causes of the probability of the occurrence of the event-effect. However, when talking about the cause of an event, often one does not intend to refer to the entire sufficient condition that is at its basis; rather, one wants to refer to one or the other of those internal components of the entire sufficient condition that are necessary to ensure that – given the further components of the entire sufficient condition – there is the production of the effect.

Expanding upon what has been said above, in the causation of y in t_1, $X \wedge x$ denotes the entire sufficient condition of event y at the previous time t_0, while x denotes a necessary component of the occurrence of y at the same time t_0. Clearly, the fact that x is a necessary condition in the production of y offers the opportunity of understanding the relation between x and y counterfactually (the way Lewis means it): if x did not occur in t_0, y could not occur in t_1. More recently, this idea has been taken up by Menzies (List & Menzies, 2009; Menzies, 2013) in the form of difference-making cause. According to him (2013, p. 78):

$$xC_{dm}y =_{def} x \Box\!\!\rightarrow y \wedge \neg x \Box\!\!\rightarrow \neg y$$

For this reason, in deploying Lewis's semantics of counterfactuals, $xC_{dm}y$ means that in all the w worlds closest to actual world, if x occurs in w then also y occurs in w and if x does not occur in w, not even y occurs in w. As emphasised by Meixner, however, the usefulness of this semantics in the analysis of mental causation depends on the way in which one conceives of the relation of closeness of the w worlds. How should we define the closest worlds to the actual

world if we claim that both $(X \wedge x)Cy$ and $xC_{dm}y$ hold? Let us remember that $(X \wedge x)Cy$ means that the conditions described by $X \wedge x$ constitute the whole sufficient condition of y, and that, in the context of the other conditions expressed by X, x is a necessary condition for y to occur. But, then, that a world is the closest to the actual world means that it is a world in which X holds.

This said, $xC_{dm}y$ means that in all the worlds in which X occur – and which are nomologically closed with respect to the actual world, if x occurs also y occurs, and if x does not occur not even y occurs. But if closeness were understood in this way, we would soon meet serious trouble. It should be noted, in fact, that X depends, in turn, on the counterfactual cause that is the subject of investigation. Now, in the context of the mental causation of a physical event P, both mental cause M and its neurophysiological correlate N, are subjects of investigation. Therefore, how should the complex of the further conditions X be understood, if the cause in question is M? X must contain, as a necessary condition, its physical correlate N. Conversely, if N is the subject of investigation, and Y is the set of the conditions that, together with N, constitute the whole sufficient condition of P, Y must also contain, as a necessary condition, the mental cause M.

Let us then ask how we should define, in terms of counterfactual semantics, the necessity of M. We should say that, in all worlds in which X (thus also N) holds, if M occurs, then P occurs, and that in all worlds in which X (thus also N) holds, if M does not occur, then not even P occurs. Conversely, in order to define the necessity of N, we should say that in all worlds in which Y (thus also M) holds, if N occurs then P occurs, and that in all worlds in which Y (thus also M) holds, if N does not occur, then not even P occurs. But, is it possible to hypothesise the existence of worlds in which N holds, and not M or M holds, and not N?

In order to give an affirmative answer, one must presuppose that N and M are nomologically independent, which is highly implausible. As can be seen from the whole debate on mental causation, the fundamental problem is precisely that of the relation between M and N. Now, M is clearly dependent on N, at least in the sense that, in the actual world, M would not occur without N. Typically, this claim is made by those who think that N plays the role of the physical realiser of M, but supporters of a weaker dependence relation would also struggle to deny it. The inverse relation of dependence is also important: N would not occur without M. Now, interactive dualism is characterised precisely by the fact that M is a different event from N and, nevertheless, N would not occur without M (at least in all the worlds that are the closest to the actual

world).[13] It follows that it is not possible to define the idea of difference-making cause, if the relation of closeness is to be understood with respect either to the conditions X, or to the conditions Y, introduced above. However, the solution is possible, if we introduce the idea of a *composite* difference-making cause. In the example in question, let $X = Z + N$ and $Y = Z + M$. That is, Z denotes the set of all conditions relevant for P to the exclusion of N and M. It is then acceptable that the following nomologically necessary implications holds:

$Z \wedge N \wedge M \Rightarrow P$

$Z \wedge \neg(N \wedge M) \Rightarrow \neg P$

(or $Z \wedge P \Rightarrow N \wedge M$)

From this it can be easily derived that:

$(Z \wedge N \wedge M) C_{suf} P$ and

$(N \wedge M) C_{dm} P$

which means that $N \wedge M$ denotes a *composite* difference-making cause. In fact, in all the worlds that are the closest to the actual world because Z holds in them, if $N \wedge M$ holds true, then P holds true and if $N \wedge M$ does not hold true, then not even P holds true. From this result, with the further assumption that M and N do not denote the same state of affair – as is obvious for every position except identity theory – we can draw the following final conclusion:

$(Z \wedge N \wedge M) C_{suf} P$ and $N C_{dm} P$ and $M C_{dm} P$.

Clearly, my perspective combines the two conceptions of cause that give contending interpretations of the relation of mental causation. On the one hand, this perspective moves from the presupposition that an actual event cannot be causally explained if not by the entire condition that is actually sufficient to determine the occurrence of the event. On the other, my perspective shares the idea of the counterfactual cause, according to which a cause is necessary, in the context of the whole sufficient condition, on pain of exclusion of the effect.

Since, in the formulation of my proposal, I have explicitly referred to Men-

[13] Note that this also applies to non-reductive physicalism, but in a much softer way. N could not be without M, but only because M formally characterises all its possible realisers.

zies, I have to say something about what divides my way of thinking about the notion of difference-making cause from his. There are two reasons for this difference. First, my notion of difference-making cause belongs to the context of sufficient causation. Second, the relation of closeness among worlds (on which the notion of necessary counterfactual cause lies) is fixed, being determined by the actual conditions *Z*, and by the laws of the actual world. For Menzies, the conditions of closeness are not subject to specific constraints. The proof of this is the fact that he does not say whether closeness involves the identity of the conditions that, in addition to *N* and *M* occur in the actual world. Indeed, one can safely say that they cannot be the same, because it is possible that, in the closest worlds, *M* acts through a different physical realiser, and this would be ruled out by *Z*, which, among the relevant actual facts, excludes the presence of realisers of *M* aside from *N*.

4.3 What Relation between Mental and Physical Causes?

Of course, the two key features of my model – collectively sufficient causes, and counterfactually necessary components – are very important in excluding the phenomenon of overdetermination. In particular, that mental causation rests upon necessity allows us to avoid the risk of overdetermination – even in some weak form – that is found in Meixner and Lowe's models. Let us remember that the phenomenon of overdetermination poses some problems from the point of view of causal attribution. In our case, if the causation of *P* on the part of *M* and *N* were overdetermined, one of the two would be irrelevant, which would mean that *P* could occur, for example, without *M*. But that is what interactive dualism should avoid. For an interactionist, the efficacy of the mental cause should be indispensable. Actually, the mixed interactive model (MIM) that I intend to propose is formally analogous to Meixner's theory of representation. Even for MIM, every mental event corresponds to a neural event. But, unlike Meixner, I conceive this correspondence to be a unique mixed neuro-mental event, characterised by two essential components, the mental *M* and the neurophysiological *N*. On this picture, it is obvious that the causal explanation of *P* requires the presence not only of *N*, but also of *M*, since *P* cannot be explained by one without the other, but only by both together. On the basis of the formal analogy between MIM and Meixner's interactive parallelism, my model also obeys a weak principle of causal closure. Its form is structurally similar to Meixner's, but has a very different meaning. This difference depends on the fact that – in-

stead of the relation of sufficient causation, xCy – my principle of causal closure contains the relation of difference-making causation:

$$\forall y[Fy \wedge \exists z(Mz \wedge zC_{dm}y) \rightarrow \exists x(Fx \wedge xC_{dm}y)]$$

For this reason, the resultant principle is a principle only of partial closure, since it states that, in the case of voluntary physical behaviour, such as the arm's movement, chains of causal dependence come to meet, amongst which at least one is constituted by physical causes only. I believe that no real form of interactive dualism can satisfy a principle of total physical closure, and that, consequently, also the principle of completeness of physics should be rejected. From the acceptance of a principle of partial causal closure, however, does not follow that mental causation can be detected by means of physical observations. Physics, in fact, can only go backwards along the series of the physical conditions on which a certain phenomenon depends. However, one cannot draw from this the conclusion that those causes are also sufficient, because they are such only together with mental causes. Someone could argue that physics is able to verify the insufficiency of the physical causes in explaining a certain event. Let us try to imagine how this could happen. The observation of neuro-physiological processes would show that there are gaps in the causal chains. The prior conditions would not be sufficient to explain the set of the subsequent effects, but possibly only their probabilities. Among the links of the incomplete chain of necessary causes, it would open a space for the intervention of the mental cause. Now, all of this could well be true, but the important fact is that it cannot be affirmed on the basis of reasons of a physical character. In conclusion, psycho-physical interaction is outside of the scope of physics, invisible to it.

The reflections just made allow us to expand further upon Lowe's conception, and to compare it to MIM. Lowe also claims that mental causation is invisible to empirical observation, but the reason put forward for affirming this is different. He does not believe that invisibility owes to the gaps that characterise the chain of physical causes, but rather to the mental cause being weakly overdetermined with respect to physical causes. Let us remember the distinction between strong and weak overdetermination. Strong overdetermination is when there are two causes, which are both sufficient to produce the same effect (for example the case of the two murderers killing the same person at the same time). In this case, no relation exists between the two sufficient causes: they are completely independent. Weak overdetermination is, instead, when causes

are at stake that are still sufficient, but are dependent on each other. This is, for example, the case when one cause is a part or a component of another. Both of them act, but they are not both necessary, such as in the case of a rotating mechanism, which uses two gears to rotate a wheel, when one would be sufficient. Well, according to Lowe, the mental cause's action yields a situation of overdetermination of this type, where the final event could also be considered the effect of a chain of purely physical sufficient causes (Lowe, 2008, p. 57). This is why mental causes will never be able to be grasped through physical observation. Unlike MIM's way of seeing things, physical observation won't even grasp any gaps in the chain of physical causes, because these, given overdetermination, cannot appear, as they do not exist.

In spite of overdetermination, Lowe attributes a precise function to the mental cause. By virtue of its intentional nature, it allows us to explain the fact that the chains of physical causes converge in the realisation of the final effect (for example, the arm's movement). Now, how can we interpret Lowe's inspiring thought, and apply it to MIM's framework? Since, within MIM, mental causes are difference-making, what function may be attributable to them by virtue of their intentional nature? They play the role of genuine final causes. As it is known, final causes are ruled out by science, because they cannot play an explanatory function. They cannot do this, because explaining means providing something that is able to produce the effect. But what could be the efficient power of a final cause? Nothing, because the *telos* for which something has been made does not influence the fact that the *telos* is reached. If the object that is endowed with a *telos* also reaches it, this happens in virtue of a cause that is different from the final one. This cause can well be the *intention* to fulfil the *telos*. In other words, a *telos* has productive power only if it is intended. Actually, an intended *telos* is a mental cause. And how, according to MIM, does this *telos* act? The answer is: by acting on the neuro-physical component of the causal chain leading to bodily movement. Note, the mental cause does not trigger a neurophysiological causal chain of which it is the first link, but, as in Lowe's view, it plays a holistic function, enabling – within the context of suitable neurological conditions – a host of neural ramifications, leading – according to plan – from the *telos*, to action. The difference between this and Lowe's view is shown by the fact that the mental cause contributes in a decisive way – together with the neurological causes – to the *production* of the causal effect.

Let us now examine Gibb's double prevention model. From the viewpoint of the theory of causation adopted by Gibb, the double prevention model is char-

acterised by the claim that not all of the physical effects have a set of causes that collectively constitute a sufficient condition for them to occur. As we know, this happens in virtue of the fact that certain events cannot be explained through conditions sufficient to their production, but only through conditions capable of preventing their prevention. These conditions serve the function of enabling the effect, not of being efficient, because the effect cannot be produced by the lack of a cause capable of preventing it. The lack of a cause cannot produce anything, for the very reason that it is, itself, nothing.

This thesis is clearly incompatible with the theory of sufficient causation, in which the cause is defined as the set of the conditions that are collectively sufficient for the effect. Among these conditions, in fact, also negative facts can – and, indeed, must – be included. Why then should these facts not have powers? Note that a negative fact does not correspond with a void of being, and thus, of power, but with an alternative state of affairs endowed with its powers. For example, the lack of an arm leaves the body without the power of the missing arm, but, on the other hand, induces a modification in the behaviour of the person who lost the arm. Is the latter perhaps not as real a power as the arm's?

A further difficulty with the model concerns the very notion of power. Let us come back to the illustration of the third scenario. Remember that this scenario has been outlined to avoid the case in which m_2 – which denotes the mental state consisting in the desire of keeping the arm at rest – could prevent n_2 from activation. The action of m_2 on the neural chain has been avoided, because, *de facto*, it has not occurred. However, even if m_2 is prevented from manifesting its power, it nevertheless possesses that power. Now, it makes no sense to speak of a power that *in principle* cannot be exerted. A power can exist even if, *de facto*, it is never exerted, but one cannot assert the existence of a power if this is not *metaphysically exercisable*. This would be like saying that one can produce something metaphysically impossible. Rather, one of two things: either this power exists, and thus it would be correct to say that a mental cause can directly act on the chain of physical causes; or this power does not exist and, then, there is no reason to interpret the mental event as an enabling condition, though not as a cause, of the arm's movement.

5. Conclusion

Nearing the end of this essay, I would like to make a few critical remarks about the kind of emergent dualism that both Lowe and Gibb endorse. The emergentist side of this kind of dualism lies in the idea – shared by physicalists – that mental events are the products of sufficient physical causes. From dualism, instead, this kind of emergentism borrows the idea of a genuine and autonomous causal role for mental states in the genesis of intended physical behaviour. «According to emergentism, non-physical mental states are causally autonomous and yet are themselves ultimately the products of prior physical evolution» (Lowe, 2008, p. 41). My question is now: are claim 1 – that mental events are the products of sufficient physical causes – and claim 2 – that mental causation is a genuine and autonomous power – really mutually compatible?

First, let us see how Lowe deals with the autonomy of mental causation. The author's crucial distinction between fact and event causation puts an emphasis, on the one hand, on the specific role of mental causes, but, on the other, imposes constraints – of an emergentist sort – to their application to the physical domain. Indeed, the role of the mental cause as a unifying factor towards myriad interwoven physical causal chains is, as we know, the mark of the mental. But this function, fulfilled beyond the transitive causal closure of the physical, does not correspond with an equally relevant role of mental causes within the causal closure. In this context, they do not seem to be on an equal footing with physical causes. Rather, mental causes often seem to have an auxiliary function to the physical: they help physical causal chains to unfold, being both the effects of physical causes and the causes of physical effects (Lowe, 2008, p. 73). Thus, the emergentist component of Lowe's dualism makes it quite difficult to guarantee the full causal autonomy of the mental, and to conceive of the mental causes as full-blown causes.

Gibb takes up Lowe's emergent dualism and radicalises it: while physical causes exert their powers on mental causes, they do not have causal powers to exert on the physical, because they are only enabling conditions. As Gibb's view touches epiphenomenalism, one might wonder whether her position is really a kind of interactive dualism.

How, then, is Gibb's assertion to understand, that the dependence of the mental on the physical is a *causal* dependence? (Gibb, 2015, this issue). Does perhaps causal dependence go along with ontological independence? (on this possibility, see Lowe, 2008, p. 70). If so, ontological dualism implies the equal

fundamentality of the mental and of the physical. But, then, is autonomous causation not a sign of ontological fundamentality? Can we split the causal level from the ontological? The debate is open. It is important to stress that, without having taken a decision about the ontology we want, no position on mental causation can be appropriate.

REFERENCES

Broad, C. (1925). *The Mind and Its Place in Nature*. London: Routledge & Kegan Paul.

Gibb, S.C. (2010). Closure Principles and the Laws of Conservation of Energy and Momentum. *Dialectica*, 64(3), 363–384.

Gibb, S.C. (2013). Mental Causation and Double Prevention. In S.C. Gibb, E.J. Lowe, & R.D. Ingthorson (Eds.), *Mental Causation and Ontology*. Oxford: Oxford University Press, 193–214.

Gibb, S.C. (2014). Mental Causation. *Analysis Reviews*, 74(2), 327–338.

Gibb, S.C. (2015). The Causal Closure Principle. *The Philosophical Quarterly*, 65(261), 626–647.

Lewis, D. (1986). Causation (with postscripts). In *Philosophical Papers*, II. Oxford: Oxford University Press, 159–213.

List, C., & Menzies, P. (2009). Nonreductive Physicalism and the Limits of the Exclusion Principle. *The Journal of Philosophy*, 106(9), 475–502.

Lowe, E.J. (2000). Causal Closure Principles and Emergentism. *Philosophy*, 75, 571–585.

Lowe, E.J. (2008). *Personal Agency. The Metaphysics of Mind and Action*. Oxford: Oxford University Press.

Lowe, E.J. (2013). Substance Causation, Powers, and Human Agency. In S.C. Gibb, E.J. Lowe, & R.D. Ingthorson (Eds.), *Mental Causation and Ontology*. Oxford: Oxford University Press, 152–172.

Mackie, J. (1965). Causes and Conditions. *American Philosophical Quarterly*, 2(4), 245–264.

Mackie, J. (1974). *The Cement of the Universe. A Study of Causation*. Oxford: Clarendon Press.

Meixner, U. (2001). *Theorie der Kausalität. Ein Leitfaden zum Kausalbegriff in zwei Teilen*. Paderborn: mentis.

Meixner, U. (2004a). *The Two Sides of Being. A Reassessment of Psycho-Physical Dualism*. Paderborn: mentis.

Meixner, U. (2004b). Causation in a New Old Key. *Studia Logica*, 76, 343–383.

Meixner, U. (2006). Causalità, teoria della. In *Enciclopedia Filosofica.* Milano: Bompiani, 1772–1783.

Menzies, P. (2013). Mental Causation in the Physical World. In S.C. Gibb, E.J. Lowe, & R.D. Ingthorson (Eds.), *Mental Causation and Ontology*. Oxford: Oxford University Press, 58–87.

Mumford, S., & Anjum, R.L. (2009). Double Prevention and Powers. *Journal of Critical Realism*, 8(3), 277–293.

Papineau, D. (2000). The Rise of Physicalism. In M.W.F. Stone, & J. Wolff (Eds.), *The Proper Ambition of Science*. New York: Routledge, 174–208.

Papineau, D. (2002). *Thinking About Consciousness*. Oxford: Clarendon Press.