Aims and Scopes

Humana.Mente is a biannaual journal focusing on contemporary issues in analytic philosophy broadly understood. HM publishes scholarly papers which explore significant theoretical developments within and across such specific sub-areas as: (1) epistemology, methodology, and philosophy of science; (2) Philosophy of mind and cognitive sciences; (3) Logics and philosophy of language (4) Normative ethics and metaethics. HM publishes special editions devoted to a concentrated effort to investigate important topics in a particular area of philosophy.

ISSN: 1972-1293

Open access statement

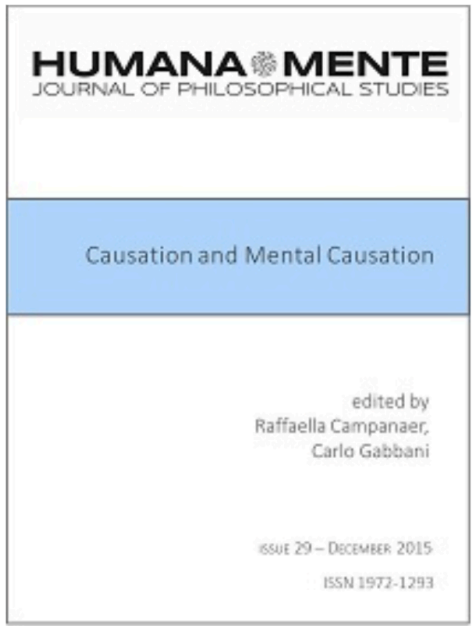
Humana.Mente adopts the principle of open access. Public access to articles is free of charge, and allows readers to read, download, copy, distribute, print, search, or link to the full texts of its articles and to use them for any other lawful purpose available under the Creative Commons Attribution-NonCommercial-NoDerivs 4.0 license.

Publishing on Humana.Mente is free of charge, the only criterion adopted for selecting manuscripts is their scientific quality.

Authorship

Papers should only be submitted for consideration once consent is given by all contributing authors. All submitting papers should be carefully checked that all those whose work contributed to the paper are acknowledged as contributing authors. This concerns all those who made a substantial contribution to the concept or design of the work. When a large, multicentre group has conducted the work, the group should identify the individuals who accept direct responsibility for the manuscript.

Peer review



edited by Raffaella Campanaer, Carlo Gabbani

It has been suggested that our conundrum concerning the possibility of mind affecting the physical world has been strongly influenced, among other things, by metaphysical choices such as considering the physical and the mental two different kinds of substance (albeit connected and interacting), or assuming a model for physical causality based on material contact, a model that is not plausible for the res cogitans. This issue of Humana.Mente aims to support and stimulate interaction and exchange between the philosophy of causality and the research directly or indirectly dealing with mental causation, presenting a wide range of reflections and possible orientations.

Published: 2015-12-01

The End of the World?
Mental Causation, Explanation and Metaphysics

Michele Di Francesco†
michele.difrancesco@iusspavia.it

Alfredo Tomasetta‡
alfredo.tomasetta@iusspavia.it

ABSTRACT

In this paper we offer some ideas on the relationship between metaphysics of causation and common explanatory practices of behaviour. We first suggest a sort of "negotiating model" for theorizing about mental causation, and then examine the so-called causal closure argument focusing on some morals one can draw from it that further illustrate the model we recommend.

Keywords: explanatory practices. metaphysics, negotiation model, mental causation, causal closure

> If it isn't literally true that my wanting is causally responsible for my reaching, and my itching is causally responsible for my scratching, and my believing is causally responsible for my saying [...] [I]f none of that is literally true, then practically everything I believe about anything is false and it's the end of the world. (Fodor, 1989)

1. Introduction

Does what we think affect what we do? As Baker (1993, p. 75) argues, apparently the answer to this question should obviously be positive. Common sense, intentional psychology, and most of the human and social sciences assume the existence of mental causation. Let us consider, for example, the explanation of the behaviour of a man who rushes to the store due to his belief that

† NEtS, Institute for Advanced Study, IUSS Pavia, Italy.
‡ NEtS, Institute for Advanced Study, IUSS Pavia, Italy.

his child will suffer without a certain medicine, and of his decision not to wait for the doctor (Burge, 1993, p. 115). It is quite natural to explain the man's actions in terms of his belief and his decision. Of course the man's rushing has both a behavioural *and* a physical explanation (the body movements are caused by physical events, after all), but there are no reasons to think that this fact would prevent the psychological explanation from being a true explanation of the man's rushing. Or are there?

Naturalistically-oriented reading of contemporary (sub-personal) science of the mind may cast doubts, and sometimes it really does, on the explicative necessity and the metaphysical solidity of the mental realm. If we could reduce the mental to the physical then we could preserve the causal efficacy of mental states: if mental properties can be reduced to physical (neural) properties, nothing prevents them acting within the boundary of the physical world. And some metaphysical "thesis", such as the causal closure of the physical realm or the denial of systematic causal overdetermination have been invoked to ground this kind of reduction: if mental states were not reducible to physical properties, then their alleged physical effects would have a physical cause – in virtue of causal closure – *and* a different mental cause, contradicting the non-overdetermination principle.

And yet, if mental states are not reducible to physical states after all – as many have thought – then, it seems, they are to be viewed as merely epiphenomenal, and our explanatory practices are therefore dramatically wrong: since a good explanation of behaviour should mention *true* causes, and the presence of physical explanations may prevent mental explanation from referring to true causes, the whole body of our explicative practices is in danger.

Now, normally we are prepared to force common sense to make room for scientific explanation, but in this case things look different – and threatening. As Fodor says in the passage quoted at the outset, practically everything that we believe about anything would be false, and this, if not the end of the world, would certainly be the end of our understanding of the world.

In what follows we explore in more detail some of the themes we have just touched upon. In particular section 2 offers some ideas on the relationship between the metaphysics of causation and common explanatory practices of behaviour, suggesting a sort of "negotiating model" for theorising about mental causation. In section 3 we focus on the so-called causal closure argument, drawing from it, in section 4, some morals that illustrate the model we recommend.

2. From Causality to Explanation?

2.1 Explanation as Prior to Metaphysics? A First Statement

The problem of mental causation originates from the difficulty of reconciling in a coherent picture some metaphysical assumptions entrenched in contemporary naturalism (such as the thesis of the causal closure of the physical, or the non-overdetermination thesis), and the pervasiveness in our explanatory practices (in science and everyday life) of the idea of the causal efficacy of the mental. A naturalistic metaphysical background seems to require that the only way we may acknowledge any causal efficacy of mental properties is by identifying them with physical properties. But there are many reasons to doubt that intentional (context-sensitive) properties and phenomenological (qualitative) properties can be reduced to physical properties. If mental states cannot be reduced to physical states then, it is often suggested, they are to be viewed as merely epiphenomenal, and our explanatory practices are therefore dramatically wrong. And since a great many of our «intellectual and practical norms» presuppose that we are agents (Burge, 1993, p. 119), and agency requires mental causation, this amounts to Fodor's dramatic claim quoted above: «practically everything I believe about anything is false and it's the end of the world».

If metaphysics threatens to lead to intellectual catastrophe, it is no surprise that some authors have proposed avoiding the "end of the world" by reversing the priority of metaphysical assumptions and explanatory practices:

> How then, are we to understand causation? My suggestion is to take as our philosophical starting point, not a metaphysical doctrine about the nature of causation or of reality, but a range of explanations that have been found worthy of acceptance (Baker, 1993, p 92).

The explanations Baker has in mind are both the scientific and the commonplace ones; her proposal is to start from explanations that "earn their keep" rather than from metaphysical assumptions (described as «freeloaders» that interfere with «the real work» of explaining and understand human action). So the proposal is to reverse the priority of explanation and causation that is favoured by the metaphysician (ibid.). This, Baker suggests, makes the problem disappear:

> If we take our ontological cue from our successful explanatory and predictive practices, then, admittedly, we end up with an ontological hodgepodge: statements concerning statutory laws, social roles, political, economic, and biological facts, as well as reasons, find their way into successful explana-

> tions and predictions. [...] My proposal, then, is to dismantle the problem of mental causation by rejecting the metaphysical background picture that generates it. If we accept paradigm cases of explanation in the sciences and in everyday life, and if we take the notion of explanation to be prior to that of causation, then the idea of a 'complete cause' in CCP [the causal closure of physics thesis] hardly makes sense (Baker, pp. 93–94).

Here adherence to the plurality of explanatory practices prevails over the search for a tidy, univocal and all-encompassing *prior* metaphysical picture, but this "explanatory stance" should not be confused with an anti-realistic view; indeed Baker is ready to acknowledge that reality does not coincide with the explainable or predictable; but – she claims – «we have no access to reality than what is required for cognitive success» (p. 95).

It is not among our tasks to discuss the details of Baker's contentious proposal (we shall say more about the relation between metaphysics and epistemology later on). However, we would like to underline a crucial point: reference to science as it is actually performed (and not to science as it should be, according to prior – questionable – metaphysical insights) may be taken as the starting point of a general argument in favour of a comprehensive view of causal explanation, sensitive both to the metaphysical and the epistemological issues involved.

In fact, one may think, the acknowledgment of the pluralistic practices of real science – and indeed the acknowledgment of their "hodgepodge nature" – is not to be confined within the boundaries of the study of mind, as the following passage from Nancy Cartwright clearly testifies:

> We live in a dappled world, a world rich in different things, with different natures, behaving in different ways. The laws that describe this world are a patchwork, not a pyramid. They do not take after the simple, elegant and abstract structure of a system of axioms and theorems. Rather they look like – and steadfastly stick to looking like – science as we know it: apportioned into disciplines, apparently arbitrarily grown up; governing different sets of properties and different levels of abstractions (Cartwright, 1999, p. 1).[1]

Cartwright's statement describes her overall view of contemporary scientific enterprise. Whoever agrees with this picture of science will consider the reductionist's appeal to scientific authority to justify her physicalist ontology[2] as

[1] See also Dupré (1993).

[2] Which, by the way, is an example of a derivation of metaphysical conclusions from (broadly) epistemological premises.

unwarranted and in contrast with the pluralistic nature of scientific inquiry as we find it.

Cartwright's (and Baker's) theses are indeed rather extreme; if we give a literal and factual interpretation of the idea of Nature as «governed in different domains by different systems of laws not necessarily related to each other in any systematic or uniform way» (ibid., p. 31), we may lose sight of the deep theoretical need for a unified picture of the world that seems to be shared by a large continuum of philosophical positions – from (classical) reductionism to (classical) emergentism – which describe the development from physical simplicity to higher-phenomena complexity as an example of diversification within a unitary universe.

To give up this minimal requirement can be seen as an almost complete rejection of (even a minimal form of) naturalism.[3] However, our point here is that to be confident of the existence of an ontological physicalist layout of reality is not something that can be simply inferred by actual scientific practice. Neither can it be obtained by reference to science's explanatory success, since this success is based on the plurality of conceptual tools and methodological and explicative practices that characterize the different special sciences (from chemistry and biology to geology and economics).

Reference to special sciences is particularly pertinent here, since it can be argued that the same line of argument that deprives psychological properties of their causal efficacy applies to all sciences different from physics (Fodor, 1974, 1989, 1997; Baker, 1993). Finding mental properties in the same basket as – say – geological or economical properties is in fact good news for the advocates of mental causation. As Robb & Heil (2014) put it, if any property (mental or otherwise) that is multiply realizable is condemned to causal irrelevance, all the properties of special sciences will be epiphenomenal. But, they say, «we happily accept biological, or meteorological, or geological properties as causally significant despite their being distinct from their physical realizers. Why then imagine that exclusion threatens the efficacy of mental properties?» (ibid.).

Fodor's position is of particular interest from our point of view because of his analysis of the relation between the epistemological and metaphysical consequences of the failure of reduction.

According to him, if reductionism were true, we should explain the existence of special sciences only in epistemological terms. Since, in the long run,

[3] Di Francesco (2010a, p. 74).

every science should be considered reducible to physics, our need of non-physical explanations in special sciences would be due only to contingencies having to do with our limited direct access to reality: «If only physical particles weren't so small (if our brain were on the *out*side, where one can get and look at them)», then we could directly apply physics to explain those sectors of reality that we are forced to tackle by sciences such as palaeontology of psychology (Fodor, 1974, pp. 112–113).[4]

Adopting a relaxed reading of Fodor, we may say that cognitive science, and specifically psychology, as any other science, does investigate nature looking for generalities grounded on natural kinds (and so 'counterfactuals sustaining'); if 'cognitive' kinds could be (functionally) decomposed in neurological (anatomical, biochemical, physical) terms, then the only reasons to adopt cognitive descriptions of behaviour would be practical and contingent.

But what if decomposition is impossible? What if «quite different neurological structure can subserve identical psychological functions across times and across organisms» (ibid., p. 113)? Here is Fodor's answer:

> Then the existence of psychology depends not on the fact that neurons are so sadly small, but rather on the fact that neurology does not posit the natural kinds that psychology requires. I am suggesting, roughly, that there are special sciences not because of the nature of our epistemic relation to the world, but because of the way the world is put together; not all natural kinds [...] are, or correspond to, physical natural kinds (ibid.).

The existence of different independent levels of description is taken as the sign of an ontological truth:

> The world, it seems, runs in parallel, at many levels of descriptions. You may find that perplexing; you certainly aren't obliged to like it. But I do think we had all better learn to live with it (ibid., p. 162).[5]

[4] Interestingly enough, Fodor (1974, p. 113) says that in fact even if our brains were outside our skull, where they could be directly observed, without the appropriate theoretical apparatus we wouldn't know what to look for.

[5] Fodor's (first) paper on special sciences is about forty years old, but the problem it raises is still here. José Bermúdez, for example, labels something very close to it "the interface problem": «The interface problem is the problem of explaining how (if at all) commonsense (or folk psychological) explanations of mental states and behaviour interface with the explanations of cognition and mental operations given by scientific psychology, cognitive science, cognitive neuroscience, and the other levels in the hierarchy of disciplines devoted to the study of the mind/brain» (Bermúdez, 2005, p. ix).

2.2 Causation and Explanation: Negotiating Priority

As we have seen, Baker claims that the solution to the mental causation problem should be obtained at the expense of a tidy metaphysical picture, a price she is well prepared to pay, since the alternative is the meaninglessness of our understanding of ourselves and of the psychological and social world.

A similar approach is taken up by Tyler Burge. Burge sees the very existence of worries about whether we can explain behaviour in terms of causally efficacious mental properties as the symptom of a philosophical disease:

> But what interests me more is the very existence of the worries. I think that they are symptomatic of a mistaken set of philosophical priorities. Materialist metaphysics has been given more weight than it deserves. Reflection on explanatory practice has been given too little. The metaphysical grounds that support the worries are vastly less strong than the more ordinary grounds we already have for rejecting them (Burge, 1993, p. 97).

Baker and Burge call for a radical change of attitude toward the relation between causality and explanation, which is in turn connected to a new conception of the relation between ontology and epistemology: a conception that takes as its starting point the concrete analysis of our explanatory practices and not the *a priori* adherence to metaphysical principles. If we take this stance, then our metaphysical theses about causation should follow from our explicative practices, and not the other way around. In this case, a pluralistic account of explanation in science works as a constraint for any metaphysics of causality:[6]

> I think it more natural and fruitful to begin by assuming, defeasibly perhaps but firmly, that attributions of intentional mental events are central to psychological explanation both in ordinary life and in various parts of psychology. We may also assume that intentional mental events are often causes and that psychological explanation is often a form of causal explanation. Given these assumptions, the 'worry' about epiphenomenalism seems very remote. [...] None of the metaphysical considerations advanced in current discussion seem to me remotely strong enough to threaten this conclusion (ibid., pp. 60–61).

Robb & Heil (2014) nicely summarize the epistemological priority approach in the following way:

[6] Di Francesco (2010a, p. 74).

> This suggests that, rather than let a priori conceptions of causation (or properties, or causal powers) lead us to regard mental causation with suspicion, we should reason in the other direction: revise our conception of causation to fit our actual scientific beliefs and practices. If the metaphysicians were right about causation, no science would be possible beyond basic physics (biological properties, for instance, would lack causal efficacy).

The "epistemological" approach, however, can be criticised in many ways. For example, it can be argued that the argument for mental epiphenomenalism does not extend to all the special sciences; or that the causal efficacy of the properties of special sciences should not be taken for granted – being subject, as everything, to the scrutiny of metaphysics; or that instead of giving up widely shared metaphysical premises, we should investigate what aspects of the commonsense view of causation leads to difficulties.[7]

The most insidious criticism, perhaps, says that epistemological approaches are based on a misunderstanding: the problem is not *whether* mental causation is possible, but *how* it is possible. Kim (1998, p. 62) raises the point with his usual perspicacity:

> The issue is not metaphysics versus explanatory practice, as Burge would have it, nor metaphysics versus epistemology, as Baker would have it. Nor is the issue one of choosing between metaphysics and mental causation: most of us have already chosen mental causation, although as philosophers we should regard pretty much everything ultimately negotiable. The issue is how to make our metaphysics consistent with mental causation, and the choice that we need to make is between various metaphysical alternatives, not between some recondite metaphysical principle on the one hand and some cherished epistemological practice or principle on the other.

Appropriately, Kim notices that the idea of starting from our understanding of mental causation has its obscurities. What is in fact our understanding? Nowadays it is common to acknowledge the causal efficacy of the mental, but in the 1950s and 1960s the issue was widely discussed under the label "are reasons causes?" – with Wittgensteinians and common-language philosophers opting for the negative answer. According to Kim (ibid. , pp. 62–63), the change in causal *Weltanschauung* is due mainly to Davidson's theory, a theory that was influential mainly because of its metaphysical theses (such as the nomological

[7] Antony (1995), Kim (1998, pp. 61–62; pp. 78–79), Antony and Levine (1997, p. 96), Crane (2001), McLaughlin (2006), Robb and Heil (2014, pp. 27–28; pp. 44–46).

character of causation and a specific theory of events).

This allows Kim (ibid. , p. 63), to say that «much of the current debate on mental causation has stemmed from a widely shared dissatisfaction with Davidson's account», and that tackling the debate requires analysing «the whole metaphysical package» of anomalous monism and the anomalism of the mental. And in any case, he goes on, reference to explanation cannot eliminate metaphysical issues since «the only way in which I believe that we can understand the idea of causal explanation presupposes the idea that the event invoked in a causal explanation is in reality a cause of the phenomenon to be explained» (ibid. , p. 64).

These are powerful criticism indeed (along the same lines, see Crane, 2001, pp. 60–61). However, in our opinion, things are less tidy than it seems. Let us give some examples.

(1) Certainly many aspects of the current debate about (mental) causation have stemmed from worries connected with Davidson's metaphysical assumptions, but some have not. Realism about mental causation was proposed as part of the philosophical hard-core of cognitive science, which is, of course, an *empirical* research program (to use Lakatos' terminology) that dominated the second half of the XX century.[8] The very existence of a highly successful empirical research program which postulates inner mental states as the cause of behaviour is in itself an explanation of the rise of a realist approach to (causally efficacious) mental properties.[9] And a metaphysical conclusion that conflicts with such a widespread scientific practice risks being regarded as an example of philosophical *hubris* (and this should be particularly worrisome for a physicalist philosopher).

(2) This in turn shows that there is a profound difference between the fifties and sixties of the XX century and today: what is at stake today is not (only) commonsense explanation of action vs. rigorous (analytic) metaphysics, but (also) the relation between a science based on mentalistic explanation and reductive metaphysics. The contrast is between a pluralistic picture of the scientific enquiry[10] and physicalist monism engendered – or at least suggested – by theses

[8] Fodor's views on the 'causal status' of special sciences exemplify this general observation.

[9] Another metaphysical consequence of the explicative success of cognitive science was the widespread adoption of non-reductive physicalism viewed as the metaphysical option most compatible with multiple realisation. Here again the direction appears to flow from explicative success to metaphysics.

[10] In recent years epistemic pluralism has been confronted by the appearance of a growing corpus of new disciplines characterized by the prefix "neuro" – neuro-economics, neuro-ethics, neuro-aesthetics, and so on. Even if a neuro-discipline explanation does not typically involve just neural facts (psychological, an-

such as the causal closure of the physical realm and the denial of systematic causal over-determination.

(3) It is possible that the proponents of the epistemological approach misrepresent the problem of mental causation (either conflating epistemology and metaphysics, or taking causation as an epistemic notion). But just saying that the problem is not *whether* mental causation is possible, but *how* it is possible, understates the relevance of their worries. The real issue is the following: could the "how-question" be completely answered in metaphysical terms, or – when engaged in the construction of a general view of the causal structure of the world – need there be room for a kind of trade-off between causation and explanation? After all, if a good explanation should mention true causes, a meta-explanation of the success of our best explicative practices may well interact with the construction of our metaphysical picture of the world.

(4) But suppose one could pursue the inquiry into the (mental) causality issue, and into the "how-question", in exclusively metaphysical terms. The fact is, we think, that this strategy is not at all advisable.

A purely metaphysical approach should indeed settle many complex issues: such as the investigation of (i) the nature of causal *relata*, (ii) properties, (iii) the nomological, (iv) substances, (v) the causal relation itself, and so on and so forth (see Gibb, 2013 for a much longer and more detailed list); and hoping to make up ones mind about these difficult problems in isolation from the concrete, and successful, explanatory practices of behaviour, is to our mind really too confident an attitude. Our suggestion is therefore a sort of "negotiation model": the construction of a metaphysical picture of the mental causation realm should always involve a continuous trade-off between metaphysical considerations on the one hand, and the suggestions offered by our successful explanatory practices on the other. This strategy should not be endorsed in virtue of its conciliatory aspect, but because it is strongly suggested by the fact that purely metaphysical principles are very far from being a stable and firm basis for our theoretical constructions. For example, many contemporary philosophers like to start from the "causal closure of the physical" principle in order to establish a physicalist ontology and to frame, within it, a more definite image of the mental

thropological and social facts are usually referred to), it is quite common to present the neural factors as the grounding elements in the explanation of the kind of behaviour under analysis. In our opinion, however, neuro-foundationalism (the claim that neural explanations are more fundamental than higher-level explanations) does not appear to be sustained by a real examination of contemporary science of the mind, that still shows the presence of a plurality of explicative styles, conceptual tools, and languages whose reduction to neuro-language is hard to envisage – see Di Francesco (2010b), Di Francesco and Marraffa (2014).

causation issue. And this kind of argumentative strategy is often believed to work because of the alleged firmness of the causal closure principle itself. But this, we think, is simply overconfident: the principle at issue is indeed a very contentious one. That the causal closure principle really is very contentious is in fact the point of the next section. We intend this discussion as a case study aiming to suggest the contentiousness of virtually all metaphysical principles involved in the causality debate – given that the causal closure principle is probably the strongest and the most easily accepted metaphysical premise that many contemporary philosophers are disposed to acknowledge. This point will then lead us again, in section four, to question the purely metaphysical approach in favour of the wiser negotiation model.

3. Causal Closure, Materialism, and Mental Causation: A Case Study

One naturally says that the scream of someone really frightened has been caused by her fear. But suppose non-reductive physicalism is true, so that fear is realized by a brain state B, which fear itself is not identical to. It is very plausible that if one traces the causal pathway backwards from the scream, one will arrive at brain state B. But if this is the case, then there no longer seems to be anything for the mental state to do: fear seems "excluded" from the causal story, contrary to our initial intuition. This is the gist of Kim's causal exclusion argument against non-reductive physicalism (Kim, 1998).

Since its appearance, Kim's argument has received much attention, generating a more and more intense and intricate discussion, often focusing on subtle metaphysical and epistemological issues. Even though many philosophers have not accepted the argument, the sheer fact of the centrality of Kim's concerns in the debate, and of his skepticism about non-reductive physicalism, has greatly contributed to give non-reductionism a bad name, or, at least, to cast a sort of preliminary suspicion on it.

To this it must be added that the growing fascination exerted by the "neuro-disciplines" on philosophers and cognitive scientists (see note 10 above) has suggested to many, more or less explicitly, a sort of "neuro-fundamentalism", according to which brains and neurons are not only very important in our understanding of the mind, but are indeed just about the only entities one should consider if one wants to pursue the study of the mind in a scientifically respectable way. The centrality of the brain seems to be underpinned by, or to express, something akin to the spirit of our age, and to many this means that there is no safe and sane intellectual alternative to a physicalist and *reductionist* theory of the mind.

And yet, even though Kim's suspicions against non-reductive physicalism, and the growing "neuro-culture" (Frazzetto & Anker, 2009), seem to pull us, in different ways, towards a strongly physicalist theory of our minds, in the scientific and philosophical literature it is not easy to find clearly stated and convincing arguments in favour of physicalism itself – granting that, if one has an argument for physicalism, then ontological materialism should be of a reductive nature. So: are there arguments for physicalism in the first place?[11] Here, as we shall see, causation – and mental causation particularly – plays a central role. But let us first consider a line of argument in favour of physicalism that has nothing to do with causation, and which is indeed a widespread one, even though very often just suggested or hinted at. The reasoning is as follows. Natural sciences have given us detailed and elegant explanations of all sorts of phenomena, success after success. These extraordinary achievements offer us, on the whole, a materialistic or physicalist picture of the world, according to which the world itself is constituted, at bottom, by tiny, mindless particles governed by a few basic laws. Now, it is true that one has no clear idea on how the mind fits in this overall picture, but – so the argument goes – one can confidently induce from natural sciences' past successes that one will eventually have a complete account of the mental within the materialistic framework drawn by the sciences themselves.

Is this inductive argument a compelling one? We do not think so. Inductions are always questionable when one considers cases that are different from those already examined and to which the evidence pertains. And there are good reasons to think that the case of the mind is indeed significantly different from, say, the cases of biological life or planetary motion. As David Chalmers, among others, has repeatedly pointed out, there seems to be an obvious difference between, for example, biological life and the case of the mind.[12] In the case of life, the *explananda* are squarely functional phenomena, such as adaptation, growth or reproduction, while a seemingly distinguishing feature of the (conscious) mind is that it involves more than function.

[11] In the present context the term "physicalism" has been used to refer to a metaphysical doctrine concerning the mind. But, if we assume with the majority of philosophers that the only non-physical (i.e. 'non-metaphysically supervenient') entities one can suspect of existing in the spatiotemporal world are mental ones, physicalism concerning the mind is no different from physicalism *tout court* (where the latter doctrine is thought of as concerning the spatiotemporal realm).

[12] See, for example, Chalmers (2003, p. 110) and (2010, p. 16). For the sake of simplicity, we limit ourselves to the case of consciousness (which is commonly agreed as the most difficult for the physicalist, but we think that similar considerations could be advanced with reference to intentionality, action and normativity.

Of course conscious states do have characteristic functions and serve some purposes, and these phenomena are in need of explanation. But "consciousness" refers to all that appears or is presented to us in experience, and these *ways of seeming to a subject* can hardly be accounted for in merely functional terms. The crucial feature of the conscious mind seems to be that the whole of conscious life appears to us in a certain way, and these phenomenological appearances pose an extremely difficult challenge to any physicalist explanation. As Chalmers (2010, p. 16) writes: «The tempting induction from these cases [i.e. the cases of the phenomena already explained by science] fails in the case of consciousness, which is not [just] a problem about physical structures and functions». So one must be aware of overconfident inductions from scientific successes to materialism: the inductive argument from science to materialism is not at all trustworthy.

Let us then consider what is often presented as the *master argument* for a physicalist ontology – and, indeed, what we consider as the only hope there is to support the general idea that some sort of physicalist metaphysics has to be true.[13] This argument is crucially concerned with the issue of causation, and particularly with the principle of the causal closure of the physical world (CC). Let us start stating the CC principle in the following way – as Papineau (2002), Robb & Heil (2014), and many others do:

(CC) All physical effects have physical causes.[14]

To see how CC may be used to support a physicalist position in the philosophy of mind, let us further consider the following three theses:

a) There exist mental states (or events).
b) If mental states exist, they are causally efficacious: in particular they have physical effects.
c) If mental states exist, they are not, in any sense, physical states.
d) The physical effects of mental states are not over-determined by distinct causes.

[13] That is to say, the only hope of supporting this general idea without embracing a particular physicalist view concerning the mind.

[14] There may be more refined formulations of the principle (see Lowe, 2003; Montero, 2003; and Garcia, 2014, for example), but the one given here is enough for the present purpose. Notice that a version of causal closure is also at work in Kim's exclusion argument (at least in its full-fledged version, as it were), so that this thesis is central in arguing both for physicalism and for a reductive version of it.

Finally consider the following combinations between some of the above theses,

(I) CC + a) + c)
(II) CC + b) + c)
(III) CC + a) + b)

and, simply assuming the truth of d), let us see what can be argued for from these combinations of theses.

Start with combination I. Given that, by a), mental states exist, it follows from c) that they are not physical states in any sense. Now suppose that they have physical effects, and let us call E one of these effects. By CC, E will have a physical cause P, which is different from M, the non-physical cause of E. So E, the physical effect of a mental state is causally over-determined, *contra* d)'s assumed truth. Therefore, one must deny the assumption and affirm that mental states have no physical effects whatsoever. On the strongly plausible assumption that they are nonetheless caused by physical states, mental states are therefore to be viewed as epiphenomenal by-products of the physical, deprived of any causal power on the material world (see, for example, Jackson, 1982, and Robinson, 2004). So, from the first combination of theses, one can infer the truth of epiphenomenalism:

(I) CC + a) + c) → Epiphenomenalism.

Let us now consider combination II, and suppose that mental states exist. From this supposition, b), and c), it follows that mental states are non-physical states that have physical effects. Let E be one of these effects. By CC, E will have a physical cause, P, which is different from M, the non-physical cause of E. So E, the physical effect of a mental state is causally over-determined, *contra* d)'s assumed truth. Therefore one must deny the assumption and affirm that mental states do not exist after all: there are brains and neurons but no mental states, and this, of course, is eliminative materialism (see Churchland, 1981):

(II) CC + b) + c) → Eliminative materialism

Given this and the previous result, one can already put together an argument for physicalism in the philosophy of mind based on the assumed truth of CC. This argument turns on the fact that epiphenomenalism and eliminativism concerning the mind, though sometimes supported by some clever philosopher or other, are nonetheless extremely implausible positions. Saying that, for example, my desiring an ice cream has nothing to do with my actually buying an ice cream seems to many too odd a position to be entertained as more than an intel-

lectual exercise. And the idea that my desires simply do not exist (and that this is the case for all other mental states), seems simply a crazy non-starter. If so, then the conjunction of a) and c), and the conjunction of b) and c) are both false. But, of course, if eliminativism is to be rejected, then a) is true, and so c) has to be false; and, analogously, if epiphenomenalism is to be rejected (along with eliminativism), then b) has to be true and so, again, c) has to be false. But denying c) means affirming that every mental state is a physical state, and physicalism is therefore vindicated. And if, moreover, one assumes, as we are doing here for the sake of argument, that the prospects of non-reductive physicalism are dim, we have an argument for *reductive* physicalism.

The point just made may be seen more straightforwardly considering combination III. Given that, by a), mental states exist, it follows, from b), that they have physical effects: let us call E one of these effects, and suppose that mental states are non-physical. By CC, E will have a physical cause P, which is different from M, the non-physical cause of E. So E, the physical effect of a mental state, is causally over-determined, *contra* d)'s assumed truth. Therefore one must deny the assumption, and affirm that mental states are indeed physical states. So some forms of physicalism concerning the mind must be true (and, again, assuming that the prospects of non-reductive physicalism are dim, we have an argument for *reductive* physicalism):

(III) CC + a) + b) → (Reductive) Physicalism

How should one evaluate the argument for (reductive) physicalism we have just seen? Thesis a) is almost unquestionable, we said; and thesis b) is just about on the same footing. So it is CC, the principle of causal closure of the physical realm, that carries the weight of the argument, and this is indeed why the argument just presented, commonly viewed as the master argument for physicalism, is known as "the argument from causal closure"[15]: one can almost say, if CC is true, then (reductive) physicalism is also true.

But is CC true?

To begin with, one thing worth noticing is that CC is not a law or a principle one can find in physics textbooks, or, for that matter, in any science textbook.[16] One may say that it is a *methodological* principle practising scientists adhere to when confronted with some physical phenomena in need of explanation, but of

[15] Note, by the way, that our way of presenting the argument is somewhat different from the usual presentations.

[16] Robb and Heil (2014).

course this methodological attitude by itself says nothing at all about the status of CC, not constituting epistemic evidence for its truth.

So are there convincing justifications for CC? Well, it is sometimes heard that CC can be vindicated on the basis of the principle of conservation of energy: this scientific principle, it is said, prohibits a non-physical entity from acting in the physical world, because otherwise the quantity of energy in the physical world itself would change, and this cannot happen.[17]

But this is not a good argument. What science has established is that the internal energy of an *isolated* system remains constant: the internal energy of an isolated system cannot be changed by any process taking place *within* the system. And this, of course, cannot rule out the possibility that some non-physical entity, not belonging to the physical world, may exert its own causal, and *external*, influence on the physical world itself.

Are there more compelling arguments for CC in the philosophical literature than the one just considered? David Papineau (2000, 2002) has put forward two, and we think that these two arguments, even though questionable, are, at the end of the day, the only ones that may be considered as worth examining.

Let us consider the first argument, the one "from fundamental forces". Papineau (2000, pp. 198–199) writes:

> 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.

What Papineau is saying is that, given the astounding success of natural sciences in reducing apparently high-level forces to a few basic physical forces, one can confidently induce that this kind of reduction will sooner or later concern all alleged mental forces or causes, showing how they belong to the physical realm. And if so, there is no room for purely mental items existing outside the physical world and causally bringing about physical effects in it: the CC principle is therefore vindicated.

But is this argument a compelling one? It does not seem so. The fact is that the argument relies on an induction from the past successes of natural sciences

[17] See, for example, Dennett (1989, p. 35).

to a future success in reducing mental forces or causes. But, as we have said above, inductions are always questionable when one considers cases that are different from those already examined, and it is far from clear, at the very least, that the case of mind, and of mental forces or causes, is significantly analogous to that of, say, muscles and muscular forces.

So let us now turn to the second argument offered by Papineau (2000, 2008), which is the so-called 'no causal gaps argument'[18].

Let us assume, for a moment, the actual existence of non-physical mental causes acting on the physical world. Now, this causal action, as empirical findings strongly suggest, should exert itself in very specific places, namely on human and animal brains; but if this is so, then the causal processes in the brain should in many ways be "gappy", due to the existence and causal efficacy of purely mental items on the firing of the brain's neurons. Yet, after a century or so of increasingly detailed studies on the brain's behaviour, we do not have any evidence that such anomalies actually take place: the brain really does seem to act on the basis of solely physical forces and causes. Therefore, so the argument goes, the existence of non-physical mental items exerting causal action on the physical world has to be rejected, and one can affirm that every physical effect must have a physical cause.

Is the no-gaps argument a good one? We do not think it is, and our basic reason for saying so is nicely and briefly expressed in this passage by Garcia (2014, p. 103):

> Physicists often emphasize how little we know about (say) deep space. Nevertheless, there are physical laws [...] that are sufficiently well established to justify their application to deep space. Similarly, neuroscientists often emphasize how little we know about the brain. However, in this case it is doubtful that there are principles of neuroscience that are sufficiently well established to justify their application to the unknown forces

What Garcia is pointing out is that, at present, the intricacies of our brains are simply not sufficiently understood: brains are too complex items for current neuroscience, and this complexity opens up plenty of space for non-physical properties having causal influence on the physical world. The fact is that we are currently too ignorant about the brain to rule out the existence of causal gaps (in the physical chain of events) due to the causal action on it of non-physical mental

[18] For this label see McLaughlin (1994).

states: as things stand (and indeed it is not at all obvious that one day things will change significantly) the no-gaps argument does not guarantee the truth of the CC principle

Let us sum up. In current philosophical literature, the causal closure argument is commonly considered as the master argument for physicalism, and this argument crucially depends on the justifiability of the CC principle. Yet it appears that the standard strategies invoked to justify this principle are very far from being satisfying – and indeed we think there are no better strategies to be invoked. But, of course, if the CC principle has a very weak status, the causal closure argument for physicalism fails.

This said, we think that there is a further and crucial point to be noted.

Suppose, contrary to fact, that the argument from fundamental forces and the no-gaps argument succeed. Then what they show in the first place is that *physicalism is true*, and the truth of CC is just a by-product of this main conclusion. The first argument, in fact, concludes that there is no room for purely mental items existing outside the material world, and the second allows for rejecting the existence of non-physical mental items as well. So if one subscribes to these two arguments, then one has a *direct* route to physicalism; and this, we believe, is no accident. Actually we are strongly inclined to think that if there is an argument A for CC, then A is also an argument for physicalism, and that A is an argument for CC *because* it is an argument for physicalism. Here is a simple way to represent what we are suggesting:

(*) A → Physicalism → CC.

And if in suggesting (*) we are on the right track, then it is simply obvious that there is no point at all in arguing for physicalism from the causal closure principle.[19]

But the causal closure argument, we have seen, is commonly considered as the main, and indeed the only, argument one can successfully give in favour of physicalism as a general doctrine, and so one has to conclude that (very probably) there is no convincing argument for physicalism as a general metaphysical outlook.

Tiehen (forthcoming), for example, seems actually to adopt this idea, saying that physicalism should simply believed to be the most credible thesis, independently of any considerations from causal closure: abandoning physicalism, he

[19] A similar attitude can be found in two recent contributions due to BonJour (2010), an anti-physicalist, and to Tiehen (forthcoming), a physicalist.

suggests, would be akin to believing in the Fairy Queen. And the underlying thesis – one we subscribe to – is that one should not argue from CC to physicalism, but that things go the other way around: *if physicalism is true, then causal closure is also true*; and, on the other hand, if physicalism *is not* true, neither is CC.

4. Some Morals

Let us now briefly state some morals we think one could draw from the previous discussion.

Many contemporary philosophers embark on a purely metaphysical inquiry into the nature of the mind, and its causal powers, exploiting the alleged metaphysical firmness of the CC principle. But the fact is that CC is very far from being an uncontroversial basis for grounding a physicalist outlook of the mind and of mental causation. Indeed, its truth seems to depend on a prior assumption of the very metaphysical outlook it should justify. So starting from CC to frame an ontology of the mind – and its causal powers – is not really advisable.

Yet one could think of starting to theorise from physicalism itself, simply assuming its truth as the most secure metaphysical ground. But why should one believe that physicalism is the most credible metaphysical outlook? One of the reasons often appealed to by physicalists has to do with the explanatory success of natural science. In fact, even though (as we said) natural sciences' past successes cannot offer a compelling inductive basis for affirming physicalism, one may still suspect that, after all, there is no sensible alternative to the kind of explanations they offer, and so, maybe, that there is no sensible alternative to physicalism. And from this basis one could start to construct a framework for analysing the mind and its causal powers.

This move, notice, brings back explanatory considerations into the metaphysical picture, and if this is so one is not entitled to focus just on the explanatory success of *natural* sciences[20].

The point we want to make here is that the (rather weak and somewhat vague) suggestion according to which natural science's explanatory success grounds physicalism is counterbalanced by the consideration of an element too often overlooked in debates of this kind: the causal explanations of human behaviour offered by common sense, intentional psychology and most

[20] Strictly speaking, causal closure should be attributed only to the bottom, micro-physical level of causal interaction. For the sake of argument, we extend its reach to natural sciences such as chemistry and biology, whose reduction to micro-physics is, however, far from obvious.

of the human and social sciences. And this means, among other things, that one crucial factor in the possible choice of a physicalist outlook should be the comprehensive consideration of the successful explanatory practices of our behaviour, leading to a sort of "physicalism with a human face". Let us elaborate a bit on this idea.

In our opinion, even the supporter of a physicalist conception of the mind and the spatiotemporal world aiming at a theory of (mental) causation should not ignore what common sense, intentional psychology, social science and humanities' causal explanations implicitly tell us about mental causation and about causation *tout court*. True, if the mental realm is simply a part of physical reality, then the causal interactions between the mental and the physical are just a particular case of physical causation, and they should not be treated as a problematic or otherwise peculiar phenomenon, deserving special attention and a somewhat broader conception of causality: all spatiotemporal causation is physical causation.

Yet, even in this case, the explanatory practices of common sense, intentional psychology, social sciences and the humanities have an important role to play for a metaphysical account of causation. If the mental is physical, then we know "from the inside" many examples of physical causation. And what is more, the direct availability of many examples of personal-level, and usually successful, causal explanations involving the mental should inform every sensible attempt to give an overall metaphysical account of the causal relation. If the mental is indeed physical, we do not just have direct knowledge of cases of physical causality, but we also have many trustworthy and detailed descriptions, explanations and predictions concerning personal agency, that is many descriptions, explanations and predictions concerning a significant part of the physical world. And any reasonable physicalist wanting to account for causal processes should not ignore, on pain of significant mistakes, all of these reliable causal explanations concerning the physical world. So aiming at a theory of (mental) causation should involve – even for the physicalist – a reflective equilibrium between the pure ontological side and the "epistemic side" of personal-level causal explanations.

In purely general terms, we believe that – even though pure ontology has its own rights – in constructing a theory of mental causation one cannot ignore the fact that the causal explanations offered by common sense, intentional psychology, social science and the humanities are well established and in some important cases truth conducive: metaphysical theories and consideration of the ex-

planatory practices of behaviour are in fact, and should be, closely intertwined in inquiring into the nature of the mind and its causal powers. This is, *de facto*, a too often neglected point, one to which, we think, philosophers generally, and causation theorists in particular, should pay more attention in future research. And it is in fact the negotiation model we already suggested, a model that, we hope, will now prove more convincing to the reader.

REFERENCES

Antony, L.M. (1995). I'm a Mother, I Worry. *Philosophical Issues*, 6, 160–166.

Antony, L.M., & Levine, J. (1997). Reduction With Autonomy. *Philosophical Perspectives*, 11, 83–105.

Baker, L.R. (1993). Metaphysics and Mental Causation. In J. Heil, & A. Mele, 75–96.

Bermudez, J.L. (2005). *Philosophy of Psychology*. London: Routledge.

BonJour, L. (2010). Against Materialism. In R.C. Koons, & G. Bealer (Eds.), *The Waning of Materialism*. Oxford: Oxford University Press, 3–23.

Burge, T. (1993). Mind-Body Causation and Explanatory Practice. In J. Heil, & A. Mele, 97–120.

Cartwright, N. (1999). *The Dappled World. A Study of the Boundaries of Science.* Cambridge: Cambridge University Press.

Chalmers, D.J. (2003). Consciousness and its Place in Nature. In S. Stich, & T. Warfield (Eds.), *The Blackwell Guide to Philosophy of Mind*. Oxford: Blackwell, 102–142.

Chalmers, D.J. (2010). *The Character of Consciousness*. Oxford: Oxford University Press.

Churchland, P.M. (1981). Eliminative materialism and the propositional attitudes. *The Journal of Philosophy,* 78 (2), 67–90.

Crane, T. (1995). The Mental Causation Debate. *Proceedings of the Aristotelian Society*, Supplementary Vol. 69, 211–236.

Crane, T. (2001). *Elements of Mind: An Introduction to the Philosophy of Mind.* Ox-

ford: Oxford University Press.

Dennett, D. (1989). *Consciousness Explained*. Boston, MA: Little Brown and Company.

Di Francesco, M. (2010a). Two Varieties of Causal Emergentism. In A. Corradini, & T. O'Connor (Eds.), *Emergence in Science and Philosophy*. London: Routledge, 64–77.

Di Francesco, M. (2010b). La penultima verità. Naturalismo e neurofilosofia. *Rivista di Estetica*, 44 (2), 7–28.

Di Francesco, M., & Marraffa, M. (2014). A Plea for a More Dialectical Relationship between Personal and Subpersonal Levels of Analysis. *Frontiers in Psychology*, 5 (5), 1–3.

Dupré, J. (1993). *The Disorder of Things: Metaphysical Foundations of the Dis- unity of Science*. Cambridge, MA: Harvard University Press.

Fodor, J. (1974). Special Sciences (or: The Disunity of Science as a Working Hypothesis). *Synthese*, 28 (2), 97–115.

Fodor, J. (1989). Making Mind Matter More. *Philosophical Topics*, LXV (1), 59–79 (Repr. in Fodor, J. (1990). *A Theory of Content and Other Essays*. Cambridge, MA: Bradford Book/MIT Press, 137–156).

Fodor, J. (1997), Special Sciences: Still Autonomous After All These Years. *Philosophical Perspectives*, 11, 149–163.

Frazzetto, G., & Anker S. (2009). Neuroculture. *Nature Reviews Neuroscience*, 10, 815–821.

Garcia, R. (2014). Closing in on Causal Closure. *Journal of Consciousness Studies*, 21(1-2), 96–109.

Gibb, S.G. (2013). Introduction. In S.G. Gibb, E.J. Lowe, & R.D. Ingthorsson (Eds.), *Mental Causation and Ontology*. Oxford: Oxford University Press, 1–17.

Heil, J., & Mele, A. (Eds.) (1993). *Mental Causation*. Oxford: Oxford University Press.

Jackson, F. (1982). Epiphenomenal Qualia. *Philosophical Quarterly*, 32, 127–136.

Kim, J. (1989). The Myth of Nonreductive Materialism. *Proceedings of the American Philosophical Association*, 63, 31–47.

Kim, J. (1998). *Mind in a Physical World*. Boston MA: MIT Press.

Lowe, E.J. (2003). Physical Causal Closure and the Invisibility of Mental Causation. In H.D. Heckmann, & S. Walter (Eds.), *Physicalism and Mental Causation: The Metaphysics of Mind and Action*. Exeter: Imprint Academic, 137–154.

McLaughlin, B.P. (1994). Epiphenomenalism. In S. Guttenplan (Ed.), *A Companion to the Philosophy of Mind*. Oxford: Blackwell, 277–288.

McLaughlin, B.P. (2006). Is Role-Functionalism Committed to Epiphenomenalism? *Journal of Consciousness Studies*, 13, 39–66.

Montero, B. (2003). *Varieties of Causal Closure*, in H.D. Heckmann, & S. Walter (Eds.), *Physicalism and Mental Causation: The Metaphysics of Mind and Action*. Exeter: Imprint Academic, 173–187.

Papineau, D. (2000). The rise of physicalism. In M. Stone, & J. Wolff (Eds.), *The Proper Ambition of Science*. London and New York: Routledge, 174–208.

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

Robb, D., & Heil, J. (2014). Mental Causation. In E.N. Zalta (Ed.), *Stanford Encyclopedia of Philosophy*. Spring 2014 Edition, URL = <http://plato.stanford.edu/entries/mental-causation/>.

Robinson, W.S. (2004). *Understanding Phenomenal Consciousness*. Cambridge: Cambridge University Press.

Searle, J.R. (1992). *The Rediscovery of Mind*. Cambridge, MA: MIT Press.

Tiehen, J. (forthcoming). Explaining Causal Closure. *Philosophical Studies*.