

## CONSERVATION LAWS AND INTERACTIONIST DUALISM

BY BEN WHITE

*The Exclusion Argument for physicalism maintains that since (1) every physical effect has a sufficient physical cause, and (2) cases of causal overdetermination are rare, it follows that if (3) mental events cause physical events as frequently as they seem to, then (4) mental events must be physical in nature. In defence of (1), it is sometimes said that (1) is supported if not entailed by conservation laws. Against this, I argue that conservation laws do not lend sufficient support to (1) to render its denial 'unscientific', and that those who accept (3) and deny (4) may consequently respond to the Exclusion Argument by denying (1) without thereby setting themselves at odds with current science. I also argue that conservation laws are compatible with (3) and the negation of (4), and that one can therefore accept conservation laws and (3) while denying both (1) and (4).*

**Keywords:** interactionist dualism, conservation laws, mental causation, exclusion argument.

The Exclusion Argument for physicalism<sup>1</sup> maintains that since (1) every physical effect has a sufficient physical cause, and (2) cases of causal overdetermination, wherein an effect is produced by two or more distinct, independently sufficient causes are not as frequent, systematic, or widespread as they would have to be if every physical effect of a mental cause were overdetermined, it follows that if (3) mental events cause physical events as frequently as they seem to, then (4) mental events must be physical in nature. For those who are loathe to either reject (3) or accept (4), the most common way of responding to the Exclusion Argument has been to contest (2), either by denying that causal overdetermination is as rare or random as (2) avers or else by questioning whether effects produced by distinct mental and physical causes indeed qualify as overdetermined. The alternative strategy of rejecting (1) is, on the other hand, typically viewed as simply too crazy to be taken seriously.<sup>2</sup> While

<sup>1</sup> The most vocal proponent of the Exclusion Argument in recent times has been Kim (1989a,b, 1993, 1998).

<sup>2</sup> To handle certain counterexamples that quantum phenomena seem to pose to the thesis that every physical effect has a sufficient physical cause, (1) may be reformulated so as to state that

reasons are rarely given in support of this assumption, the standard view appears to be that (1) follows or else draws strong inductive support from certain fundamental and well-established laws of physics: viz. the conservation laws of energy and momentum. It is hence thought that to deny (1) is to set oneself at odds with the prevailing scientific consensus. In this paper, I question whether this is in fact the case; i.e., whether conservation laws indeed lend sufficient support to (1) to render its denial ‘unscientific’. My aim will be to show that they do not, and that those inclined to accept (3) and deny (4) (hereafter ‘interactionist dualists’) may consequently respond to the Exclusion Argument by denying (1) without thereby controverting any basic law or principle of physics. In defending this claim, I will also attempt to show that conservation laws likewise do not present any direct threat to interactionist dualism, and that one can thus openly accept such laws while also endorsing (3) and rejecting (4). If this is correct, then it is perfectly consistent for one to accept conservation laws and the thesis that mental events cause physical effects while denying that mental events are physical and that every physical effect has a sufficient physical cause.

## I. HOW MIGHT CONSERVATION LAWS BE THOUGHT TO POSE A THREAT TO INTERACTIONIST DUALISM?

In their colloquial form, the conservation laws of energy and momentum state that in a closed system (i.e., one that exchanges no matter or energy with its surroundings, and on which no external force acts), the total quantity of energy and momentum always remains the same. The idea that these laws pose a problem for interactionist dualism goes back at least as far as Leibniz (1714/1898: §80), who objected to Descartes’ theory of mind–body interaction on the grounds that it violates the conservation of momentum (or as Leibniz (c.1691/1896: 667) called it, the conservation of ‘common progress’), and despite periodic fluctuations in its popularity, acceptance of this idea is still fairly widespread today.<sup>3</sup> Of those who hold this view, some (e.g., Papineau 2001) see the problem that conservation laws raise for interactionist dualism as lying in the support they lend to (1), whereas others (e.g., those cited in footnote 3) seem to see such laws as flatly inconsistent with any non-physical

physical causes are sufficient to fix the probability of every physical effect. For ease of expression, I will make use of the original, simpler formulation of (1) presented above, though nothing in the following discussion hangs on this.

<sup>3</sup> Proponents of this idea include Crane (2001: 48), Dennett (1991: 35), Fodor (1981: 114), Putnam (1999: 78–9), and Searle (2004: 42). Other advocates are noted by Montero (2006: 384–5) and Gibb (2010: 636fn1).

causation of physical effects. Against this, I will seek to show both that (1) does not follow from conservation laws and that such laws are moreover consistent with interactionist dualism.

Let us begin with the question of whether conservation laws can be used as the basis for a direct argument against interactionist dualism that is independent of (1). (For ease of exposition, I'll henceforth focus primarily on the law of conservation of energy (CoE). Most of the arguments and claims to be discussed can, however, be applied *mutatis mutandis* to other conservation laws as well.) A natural first step in attempting to construct such an argument might be to conjoin CoE with the additional premises that (i) the universe is a closed system, and (ii) that any change in a body's motion involves some transference of energy between the cause of the change and the body whose movement is altered. From these premises, one might then try to derive the conclusion that no change of bodily motion has a non-physical cause, and that interactionist dualism is therefore false. This yields Argument I:

#### Argument I

(CoE) Energy is conserved in any closed system.

(i) The universe is a closed system.

(ii) Any change in a body's motion involves some transference of energy between the cause of the change and the body whose movement is altered.

No change of bodily motion has a non-physical cause.

As many have noted, however, even granting premise (ii), the resulting argument against interactionist dualism is invalid, for the possibility is left open that mental events might themselves possess some form of energy, which they transfer to those bodies whose motion they affect.<sup>4</sup> So long as the total amount of energy that exists before and after these transfereces turns out to be the same, non-physical, mental events could be capable of causally affecting bodily motion without violating CoE or any of the other premises of the argument just proposed.

To exclude the possibility that non-physical, mental events might causally interact with physical events in this way, a *valid* argument against interactionist dualism based on (i), (ii), and CoE must hence contain an additional premise stating either that (iii) nothing non-physical has energy (or at least none that is capable of being transferred to any physical body) or (iv) that the physical realm constitutes a closed system (i.e., one that exchanges no matter or energy with its surroundings, and on which no external force acts). The choice between

<sup>4</sup> Hart (1988: ch.9, 10, 12) develops a fairly sophisticated form of interactionist dualism along these lines.

these two further premises gives us the following two arguments in place of the one first proposed:

#### Argument II

(CoE) Energy is conserved in any closed system.

- (i) The universe is a closed system.
- (ii) Any change in a body's motion involves some transference of energy between the cause of the change and the body whose movement is altered.
- (iii) Nothing non-physical has energy (or at least none that is capable of being transferred to any physical body).

No change of bodily motion has a non-physical cause.

#### Argument III

(CoE) Energy is conserved in any closed system.

- (i) The universe is a closed system.
- (ii) Any change in a body's motion involves some transference of energy between the cause of the change and the body whose movement is altered.
- (iv) The physical realm is a closed system.

No change of bodily motion has a non-physical cause.

While both of these arguments are perfectly valid, I do not think either is ultimately all that compelling. First, as noted by Montero (2006), once (iii) or (iv) is added to the original CoE-based argument against interactionist dualism proposed above, the appeal to CoE becomes unnecessary<sup>5</sup>, for the falsity of interactionist dualism can be deduced from (ii) and (iii), or (ii) and (iv) alone. The fact that CoE can be removed from the premises of Arguments II and III without in any way affecting their validity would seem to suggest that the real threat to interactionist dualism lies *not* in CoE, but rather in certain of the remaining premises of Arguments II and III, viz. (ii), (iii), and (iv).<sup>6</sup> This would be good news for dualists, for since these other premises are not self-evident, nor are they, like CoE, well-established, fundamental laws of physics, contesting them should be significantly easier than calling CoE itself into question.

<sup>5</sup> Note that (i) is thereby made unnecessary as well.

<sup>6</sup> See Broad (1925: 107, 109). Montero (2006: 395) takes this to show that CoE in fact has 'nothing whatsoever' to do with the defence of physicalism. This would be correct if Arguments II and III were the *only* arguments against interactionist dualism that CoE might figure into, but as will be shown below, that is not the case. Koksvik (2007b: 579), e.g., points out that an additional argument against interactionist dualism can be constructed by conjoining CoE and (iv) with the assumptions that '[i]f a non-physical mind changes a physical system, it changes its energy level', and that '[i]f the energy level of a physical system is changed by a non-physical system, energy is not conserved in the physical world'. Discussion of this argument will be postponed until Section V.

## II. CAN NON-PHYSICAL ENTITIES POSSESS ENERGY?

Starting, then, with Argument II, what reason might dualists have to accept premise (iii)—that nothing non-physical has energy? One such reason might be found in Averill and Keating's (1981: 105) remark that 'changes in the energy of non-physical things are undefined, i.e., there is no way of specifying the state of a non-physical thing in terms of the variables of physics'. Here the idea seems to be that quantities of energy are attributed to things only under some physical description of their states and relations (i.e., some description given in terms of the proprietary laws and kinds of physics); therefore, energy cannot be properly attributed to any non-physical thing, because non-physical things cannot be described in physical terms. To this the dualist might respond by arguing that there is no reason why a non-physical entity could not be ascribed a physical quantity if such an ascription were warranted by certain effects that it was found to have upon some physical system.<sup>7</sup> If, e.g., the occurrence of a certain mental event was found to correlate with changes in the energy level of the brain in which that event was realized, and during these changes the energy in the brain's surrounding physical environment was known to remain constant, then rather than immediately rejecting CoE, it would not seem unreasonable to instead attribute the quantity of energy needed to account for these changes in the brain's energy level to the mental event itself. Averill and Keating are therefore wrong to claim that 'there is no way of specifying the state of a non-physical thing in terms of the variables of physics', for such specifications could be made on the basis of such a thing's measurable effects on physical systems. The values of the relevant variables could be specified as required to account for those effects in a manner that is consistent with CoE.<sup>8</sup>

It is also worth noting that as it is unclear how something could exert a force without possessing energy, Averill and Keating's reluctance to attribute energy to non-physical states seems to stand in tension with their suggestion (discussed further below) that the mind exerts an external, non-physical force on the brain (Averill and Keating 1981: 103). If correct, this point is significant, because it entails that any dualist who maintains that the mind causes physical effects by exerting a force on physical entities must also allow that mental causes have energy, even though they are (according the dualist) non-physical.

<sup>7</sup> On the assumption that while non-physical entities might possess energy, they cannot possess mass, the energy possessed by non-physical entities would have to differ from that possessed by physical entities at least in the respect that when possessed by non-physical entities, it is not equivalent with mass.

<sup>8</sup> See Fair (1979: 229), who notes that 'the hypothesis that energy is a *conserved* quantity' has at times forced physicists to 'revise [their] definition of energy', thereby leading to the discovery of 'new forms and carriers of energy'. Couldn't the same procedure at some point lead us to treat mental events as potential carriers of energy as well?

Assuming, then, that ascriptions of energy to non-physical entities could indeed be justified along the lines suggested above, wouldn't the attribution of a physical quantity to such an entity nonetheless deprive the latter of its non-physical status? In other words, if mental events can possess physical quantities, what grounds could there be for treating such events as non-physical? Here the dualist might reply that so long as mental events exhibit certain features (e.g., qualitative and/or intentional features) that cannot be identified with, reduced to, or fully explained in terms of any physical quantities we might ascribe to them, there is no reason to view their possession of such quantities as somehow 'turning them into' physical events. For property dualists, in particular, this idea should not sound overly strange, for if the property dualist is correct in thinking that physical entities can bear non-physical properties, there would seem to be no reason (or at least no conceptual reason) why non-physical entities could not likewise be capable of possessing certain physical quantities.

### III. PAPINEAU'S ARGUMENT FROM FUNDAMENTAL FORCES

Averill and Keating's remark thus does not appear to provide a compelling argument for (iii)—the claim that nothing non-physical has energy. A more promising case for the claim might, however, be drawn from two arguments developed by Papineau (2001), which he calls, respectively, the Argument from Fundamental Forces and the Argument from Physiology. These two arguments draw upon the history of science as providing inductive support for the thesis that 'there are no special [i.e. non-physical] mental or vital forces' (Papineau 2001: 27). Since energy is partly defined in terms of force (for energy is the capacity to do work or transfer heat, and work is the application of force to a body that results in the displacement of that body in the force's direction), it follows that if there are no special, non-physical forces, then nothing non-physical has energy (or at least none that it can use to do work). Papineau's arguments can thus be equally viewed as arguments for (iii), inasmuch as (iii) is entailed by the thesis that they purport to establish.

As formulated by Papineau, the Argument from Fundamental Forces 'is that all apparently special forces characteristically *reduce* to a small stock of basic physical forces that conserve energy. Causes of macroscopic accelerations standardly turn out to be composed out of a few fundamental physical forces that operate throughout nature. So, while we ordinarily attribute certain physical effects to. . . "mental causes," we should recognize that these causes, just as all causes of physical effects, are ultimately composed of the few basic physical forces' (Papineau 2001: 27). In short, reflection on the history of science shows that physical effects that were at one point attributed to the operation of certain 'special' forces have typically been found, upon closer analysis, to be fully explainable in terms of a small number of fundamental physical forces (e.g.,

gravity, electromagnetism, and strong and weak nuclear forces). This would seem to suggest that the same will likely hold true of any physical effects that are now commonly accounted for by appeal to mental causes.

Two points should be made regarding the relation between this argument, CoE, and (1)—the claim that every physical effect has a sufficient physical cause. First, while it was noted above that CoE is not needed to deduce the falsity of interactionist dualism from (ii) and (iii), CoE might still be said to have some relevance for such an argument against dualism, inasmuch as it *does* play a role in the Argument from Fundamental Forces, which (as previously mentioned) can be viewed as an argument for (iii). Papineau (2001: 28) describes this role as consisting in a certain tension between CoE and the postulation of special forces that do not reduce to any fundamental physical forces, which is that ‘[a]n insistence on the independent existence of *sui generis* special forces inside bodies threatens to remove the reasons for believing in the conservation of energy in the first place. For there are no obvious grounds for expecting such *sui generis* forces to be conservative’. In other words, the main reason we have for accepting CoE is that the few fundamental physical forces we know of appear to obey it, and all other forces that have been ‘quantatively analysed’ thus far have turned out to reduce to these few forces. The postulation of special forces that do not reduce to these few forces would therefore undermine our justification for accepting CoE, for the simple reason that if the former forces are indeed distinct from the latter, we would have no reason to believe that they share the latter’s obedience to CoE. Consequently, any reason we have for accepting CoE may *ipso facto* be viewed as a reason for rejecting the existence of special, non-physical forces (and vice versa). The fact, then, that we have very strong grounds for believing that CoE is true can therefore be seen as lending added support to the idea that all forces are reducible to a few basic physical forces that conserve energy.

On the assumption that all physical effects can be explained as due to the action (or inaction) of forces, one can easily see how the support that CoE lends to the idea just mentioned might also seem to make it into a source of support for (1). Put simply, if this assumption is correct, then the thesis that all forces reduce to a small number of basic physical forces entails that every physical effect has a sufficient physical cause, so any support that CoE lends to the former thesis will thereby count equally in favour (1) as well. This in fact appears to be the main way in which Papineau sees CoE as figuring into the case against interactionist dualism, viz. as helping to generate the Exclusion Argument by providing evidence for (1). Rather than arguing directly from CoE to the falsity of interactionist dualism, Papineau thus instead seems to have something like the following in mind:

- (a) If there are forces that do not reduce to a small stock of basic physical forces that conserve energy, then we have no good reason to accept CoE.

- (b) We have good reason to accept CoE.
  - (c) All forces reduce to a small stock of basic physical forces that conserve energy. (From (a) and (b))
  - (d) All physical effects can be fully explained as due to the action of forces.
  - (1) Every physical effect has a sufficient physical cause. (From (c) and (d))
  - (2) Overdetermination is rare.
- ((4) or  $\sim$ (3)) Mental events are physical or mental events do not cause physical effects.

This argument I think captures the part that Papineau sees CoE as playing both in the Argument from Fundamental Forces and in the more general case against interactionist dualism.

There are at least two major criticisms that dualists can raise to the preceding arguments. The first, which has been made by Garcia (2014: 102) and Koksvik (2007a: 133–4), is that the Argument from Fundamental Forces begs the question against dualism, because the inference it makes from the successful reduction of various *physical* forces (e.g., friction) to the conclusion that all forces will ultimately be so reduced relies on the very sort of presumed similarity between mental and physical entities that the dualist is apt to deny. Since the dualist maintains that the mind is importantly different from any physical thing, she/he will naturally be sceptical of the claim that any mental forces are likely to reduce to a small stock of conservative, physical forces just because certain physical forces have been so reduced in the past. On their view, mental and physical entities are *disanalogous*, so the fact that various physical forces have been reduced to a small number of conservative, physical forces gives us no reason to expect that the same will hold true of mental forces (if such forces exist). The Argument from Fundamental Forces therefore does not make a cogent case for (iii)—the claim that nothing non-physical has energy.

The second criticism is directed at the support that CoE is alleged to lend to the Argument from Fundamental Forces and thus also to (1)—the claim that every physical effect has a sufficient physical cause. Here, again, Papineau's idea seems to be that our reasons for accepting CoE count in favour of the view that all forces reduce to a small number of fundamental physical forces in terms of which all physical effects can be explained, because we have no reason to suppose that any additional, non-physical forces would conserve energy like the known fundamental physical forces do. As he himself notes, though:

[T]his is scarcely conclusive. Those thinkers who remain convinced. . . that there must be irreducible special forces inside living bodies, could still respect the universal conservation of energy, by maintaining that these extra forces must themselves operate conservatively. In support of this they could [offer] the alternative inductive argument that, because all the *other* fundamental forces examined so far have turned out to be conservative, we should infer that any extra. . . mental fundamental forces will be conservative too (Papineau 2001: 29).

In short, the dualist might argue that the postulation of non-physical, mental forces is not at all in tension with acceptance of CoE, because the conservative nature of the basic physical forces we now know of gives us ample reason to think that any non-physical, mental forces there are will likewise obey CoE.

While this is, I think, the right thing for the dualist to say, there is nonetheless at least an apparent conflict between this argument and the reasons given above for rejecting the Argument from Fundamental Forces. How, one might ask, can the dualist cite the conservative nature of basic physical forces as evidence for the assumption that any non-physical, mental forces will likewise conserve energy, while at the same time claiming that the mind is so unlike any physical thing that the reduction of various physical forces to a few basic physical forces gives us no reason to believe that mental forces (if such there are) will eventually be so reduced as well? The former idea seems to assume what the latter denies: viz. that physical and mental forces are sufficiently similar that traits possessed by the one (e.g., obedience to CoE and/or reducibility to a few basic physical forces) can be justifiably attributed to the other. Which is it then? Are mental and physical forces similar enough to warrant such inferences, or aren't they?

This dilemma can be resolved by noting that we have much greater reason to think that all forces conserve energy than we do to believe that all forces will reduce to a small stock of basic physical ones. CoE is, after all, a well-established scientific law, whereas the latter conjecture is not a law of any science. It is, moreover, somewhat misleading to present the history of science as providing unequivocal inductive support for the reducibility of all forces to a few physical ones, since scientists have also seen fit to *add* to the stock of fundamental forces in cases where certain newly discovered interactions could not be fully accounted for in terms of the forces then viewed as basic.<sup>9</sup> Given, then, that the number of forces thought to be basic and irreducible has been increased in the past, it seems reasonable to allow that further additions to this stock could be made in the future<sup>10</sup>, and who's to say that these additions might not include certain forces that are mental rather than physical in nature?<sup>11</sup> The claim that all forces will ultimately be reduced to a few basic

<sup>9</sup> The postulation of the weak and strong nuclear forces in order to account, respectively, for the phenomenon of beta decay and the coherence of the nucleus despite the electromagnetic repulsion between protons might be cited as examples of this point.

<sup>10</sup> A similar point is made by Popper in Popper and Eccles (1977: 542–3). This idea might also be supported by recent suggestions that the postulation of additional fundamental forces besides the usual four is needed to overcome certain difficulties that the Standard Model has in accounting for dark matter and the accelerated expansion of the universe. (See Reich (2010), Battersby (2013), Feng and Trodden (2014), and Dobrescu and Lincoln (2015)).

<sup>11</sup> While I assume throughout the present paper that some substantive, non-vacuous formulation of physicalism can be given, Crane and Mellor (1990: 206) see our inability to provide any principled grounds for denying psychology the authority to make such additions to our ontology as indicating that 'there is no divide between the mental and the non-mental sufficient even to set physicalism up as a serious question'. Thanks to an anonymous referee for pointing out the need to make note of this point.

physical forces hence seems dubious enough on its own that dualists needn't commit themselves to the view that mental and physical forces are *completely* disanalogous in order to deny it. They can therefore allow that any mental forces are liable to at least be similar enough to known physical forces that the former's obedience to CoE can be plausibly inferred from that of the latter, without also having to concede that the reducibility of any mental forces to fundamental physical ones can be justifiably inferred from the fact that many physical forces have been so reduced in the past.

#### IV. PAPINEAU'S ARGUMENT FROM PHYSIOLOGY

The preceding considerations seem to show that there is no straightforward, decisive argument from CoE to either (iii)—the claim that nothing non-physical has energy—or (i)—the claim that every physical effect has a sufficient physical cause. This, however, leaves Papineau's second argument for (iii), the Argument from Physiology, unaddressed. As presented by Papineau, the Argument from Physiology '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 2001: 27). While this argument can be seen as 'operating against the background provided by the... argument from fundamental forces', inasmuch as the assumption that all physical forces reduce to a few fundamental ones makes it much easier to say what sort of changes in living bodies would count as evidence of the action of non-physical forces (*viz.* any changes that cannot be explained in terms of those few physical forces deemed basic), it does not share the latter argument's dependence on CoE (Papineau 2001: 30). Here the point is merely that when we look in those places where additional non-physical forces would be most likely to manifest themselves, we do not find any physical changes that might be attributed to their influence and thus taken as evidence of their presence, for all such changes can be fully explained in terms of the physical forces already at our disposal. The problem is thus not that the putative non-physical forces may not conserve energy, but rather that they don't seem to produce any effects.

Although the Argument from Physiology itself 'has little to do with the conservation of energy', it nevertheless demands a response from dualists who are interested in using the arguments advanced in the previous two sections to develop an account of mental causation that rejects (i)—the claim that every physical effect has a sufficient physical cause—without violating CoE (Papineau 2001: 30). To see why, recall that the basic idea behind the account proposed in Section II was that mental causes might possess certain quantities of energy that enable them to exert a non-physical force on physical entities (*viz.* parts of the brain) and thereby alter the motion of such entities in ways

that cannot be accounted for in terms of any purely physical events. Section III sought to show that such an account does not stand in any necessary conflict with CoE. The dualist, however, does not seem entitled to rest content with this result, for even granting that CoE is consistent with the postulation of non-physical, mental forces that produce physical effects for which there are no sufficient physical causes, the question then naturally arises: What grounds do we have for thinking that such forces actually exist? The Argument from Physiology holds that the fact that we have found no evidence of such forces gives us reason to believe that there are none. Unless the dualist can offer some rebuttal to this argument, the work that has been done towards establishing the consistency of CoE with the account of mental causation proposed in Section II will seem rather pointless. For nothing much follows from the observation that non-physical, mental forces *could* produce physical effects that lack sufficient physical causes without violating CoE if such forces don't exist.

While Papineau appears to view the Argument from Physiology as offering a more conclusive refutation of the existence of non-physical forces than the Argument from Fundamental Forces, the response to the former argument is actually much simpler than that given to the latter above. In this case, the dualist need only point out that while neuroscience has made astounding progress in the past century, there can be little doubt that our scientific understanding of the inner workings of the brain is currently still in its initial stages. As such, it is extremely premature to claim that all changes of acceleration that take place within the brain can be fully explained in terms of the action of physical forces. While many important neural processes (e.g., the transmission of signals across synapses, or the modification of neural pathways through learning or memory) have been analysed into more basic biochemical processes that perhaps can be explained in such terms, further research may yet uncover circumstances in which the physical forces acting on the material constituents of a brain are unable to account for certain subsequent increases/decreases in the firing rate of certain neurons or the amounts of certain neurotransmitters being released. Until our understanding of the brain develops to the point where such possibilities can be ruled out, to dismiss them seems a bit hasty, to say the least.<sup>12</sup> And although Papineau is right to point out that we have yet to find any positive evidence for the existence of non-physical forces in the brain, given, again, that neuroscience is still in its infancy, to take this as evidence *against* the existence of such forces is to argue from ignorance. The absence of such evidence at these early stages should, moreover, be expected, seeing as one must first have an understanding of what changes within a system *can* be explained by the action of physical forces before one can be in a position to recognize occurrences within the system that *cannot* be explained in such terms. It is hence only after we have acquired a more fully developed understanding

<sup>12</sup> See Garcia (2014: 103).

of the biochemical mechanisms that account for the various forms of neural activity that take place in the brain that we can expect to be able to identify changes of acceleration in the brain that cannot be attributed to the action of physical forces. For these reasons, the Argument from Physiology does not provide any compelling grounds for accepting (iii)—that nothing non-physical has energy.

## V. IS THE PHYSICAL REALM A CLOSED SYSTEM?

As none of the arguments for (iii) have stood up to scrutiny, dualists seem entitled to regard the possession of energy by mental states as an open possibility, and so long as this possibility remains open, the argument against interactionist dualism from (iii) and (ii)—the claim that all changes of bodily motion involve a transference of energy between the cause of the change and the body whose motion is altered—is at best inconclusive. The argument from (ii) and (iv)—the claim that the physical realm is a closed system—has, however, yet to be answered. Fortunately for dualists, both of these premises are contentious. With regard to (iv), Montero (2006: 386–8) maintains that in contrast to the view that energy is conserved in the universe *as a whole*, the idea that energy is conserved ‘among the physical components of the universe’ is ‘a philosophical principle rather than a law of physics,’ for ‘while physics gives us reason to believe the [former conjecture], it does not seem to give us reason to believe [the latter]’. If this is correct, then given that CoE applies only to closed systems, the alleged fact that current science only supports the claim that energy is conserved in the universe as a whole implies that the only closed system that we have reason to believe exists is the universe itself, in its entirety. To assume, therefore, that the physical realm constitutes its own closed system would be to assume either that everything in the universe is physical (i.e., that physicalism is true) or that in addition to the universe in its entirety, there is also a unique subsystem within the universe that likewise conserves energy. The former assumption begs the question against dualism, and the latter is, according to Montero, unsupported by current science. Either way, dualists would seem justified in accepting CoE while rejecting (iv), provided that Montero’s contention is well founded.

Montero, however, offers little support for her claim that the restriction of CoE to the physical realm is ‘a philosophical principle’ unsupported by current science,<sup>13</sup> and reasons can indeed be given for thinking that it is instead under such an interpretation that CoE has the most evidential support. Thus, following Koksvik (2007b: 579–80), one might argue that since ‘the experimental evidence in favor of CoE has resulted from observations of

<sup>13</sup> Thanks to an anonymous referee for pointing this out and for pressing me to provide a more thorough response to Koksvik (2007b).

entirely physical systems, . . . [t]he evidence we have for CoE is *only* evidence for the restricted version', which 'holds that energy is conserved among the physical components' of the universe. Given, moreover, that CoE only applies to closed systems, the fact (if it is a fact) that CoE applies to the physical realm would likewise imply (iv)—that the physical realm is a closed system. If this argument for (iv) is successful, then in addition to the argument from (iv) and (ii), interactionist dualism will also, as Koksvik (2007b: 579) notes, be threatened by the following argument from (iv) and CoE:

Argument IV

(CoE) Energy is conserved in any closed system.

(iv) The physical realm is a closed system.

(v) 'If a non-physical mind changes a physical system, it changes its energy level'.

(vi) 'If the energy level of a physical system is changed by a non-physical system, energy is not conserved in the physical [realm]'.

'It is not the case that a non-physical mind changes a physical system.'

In light of these challenges, the prospects for interactionist dualism would be greatly improved if grounds could be given for doubting (iv) and the corresponding claim that energy is conserved in the physical realm.

To provide such grounds, dualists might start by suggesting that the only reason that current evidence seems to support the restriction of CoE to the physical realm is that we have yet to fully map out all the various exchanges of energy that take place in those circumstances in which the restricted CoE is most likely to be violated, viz. those wherein a certain brain state gives rise to or immediately follows the occurrence of some mental event. As long as this remains the case, the restricted CoE must be regarded as at best underdetermined, for a hypothesis can hardly be regarded as empirically confirmed if it has not been tested in the very circumstances where it stands the greatest risk of being falsified. Until such tests have been carried out, the dualist thus seems justified in holding that if Koksvik (2007b: 580) is correct in claiming that '[t]he evidence we have for CoE is *only* evidence for the restricted version', then the evidence for CoE is consequently much less conclusive than is typically thought. For the restricted formulation of CoE could very well be falsified when tested in those conditions where its failure seems most likely, thereby requiring us to either expand the scope of CoE and postulate certain non-physical entities that exchange energy with physical entities in such a way that energy is conserved in the universe as a whole, or else allow that energy is not conserved after all. At any rate, since current evidence does not exclude the possibility that violations of the restricted version of CoE may be discovered in the brains of minded beings, and it is this possibility that seems to pose the greatest threat to the idea that energy is conserved in the physical realm,

current evidence seems insufficient to justify acceptance of CoE as restricted to the physical realm. If this is so, then any attempt to infer (iv) from the restricted formulation of CoE will be equally suspect.

In response, one could argue that while we are at this point still unable to determine whether or not violations of the restricted version of CoE occur in the brain, we nevertheless have enough evidence indicating that other kinds of physical systems do not undergo any losses or gains of energy that cannot be accounted for in terms of compensatory exchanges of energy with their physical surroundings to enable us to infer that energy is conserved in the physical realm as a whole. Even if the restricted CoE has yet to be tested in the conditions under which it stands the greatest risk of being falsified, the wealth of other evidence in its favour might thus be thought to give us sufficient reason to accept it as true. Against this proposal, however, dualists can point out that since, on their view, the mind is (in certain significant respects) *unlike* any physical thing, those physical systems (*viz.* brains) that *interact* with minds should consequently be expected to behave in ways that other physical systems do not. It is, indeed, for this very reason that any evidence falsifying the restriction of CoE to the physical realm seems most likely to be found in the brains of minded beings, rather than in other types of physical systems; *viz.* because unlike other types of physical systems, brains (from the dualist's viewpoint) are capable of directly affecting and being affected by non-physical minds. If this is so, however, then brains are themselves importantly *disanalogous* with other types of physical systems, so the fact that other types of physical systems behave in a manner that is consistent with the restricted version of CoE gives us little reason to assume that the same will prove true of brains as well. To assume that brains are similar enough to other types of physical systems to enable us to infer that brains (and indeed *all* physical systems) behave in a manner consistent with the restricted CoE simply because other types of physical systems have been found to so do is thus to beg the question against the interactionist dualist, who is apt to deny the assumption on which this inference rests. In sum, until our understanding of the brain reaches the point where we can rule out the possibility that the energy levels of the brains of minded beings can change in ways that are inconsistent with the assumption that energy is conserved in the physical realm, dualists seem entitled to maintain that our current evidence is inadequate to justify belief in the restricted version of CoE or the corresponding assumption that the physical realm is a closed system.

## VI. DOES ALL CAUSATION OF BODILY MOTION INVOLVE TRANSFERENCE OF ENERGY?

If the foregoing considerations succeed in showing that (iii) and (iv) are at least contestable, the dualist of course has no need to also call (ii)—the claim

that all changes of bodily motion involve a transference of energy between the cause of the change and the body whose motion is altered—into question in order to defuse the arguments against their position from (ii) and (iii), and (ii) and (iv).<sup>14</sup> Nevertheless, it is at least worth noting that (ii) is rather dubious as well. The idea that any change in a body's motion requires some transference of energy between that body and the cause of the change is naturally viewed as an expression of a more general transference theory of causation of the sort advocated by Fair (1979), Salmon (1997), and Dowe (2000: ch.5, 7), according to which causation is itself simply the transference of some conserved quantity (e.g., energy) from one thing to another. One of the more common objections to such theories is that the analysis of causation they propose seems applicable only to concrete physical entities, and is hence unduly narrow.<sup>15</sup> In particular, it is difficult to see how transference theories can accommodate instances of causation by omission, since absences (being non-existent) cannot have or transfer conserved quantities.<sup>16</sup> Assuming, therefore, that it could in certain circumstances be truly said that the death of a plant was caused by the gardener's failure to water it, or that a person's religious experience was caused by a lack of food, such cases would appear to constitute straightforward counterexamples to any transference account of causation.<sup>17</sup> The same objection can likewise be applied directly to (ii), for the motion of a body would likely be significantly altered if it were to enter a void.<sup>18</sup> Due to the absence of surrounding air pressure, a human body placed in such conditions would quickly burst into bits. This would certainly constitute a change in the body's motion, and the surrounding void seems clearly to be the cause of this change, yet since the void does not exist, the effect it has on the body's motion cannot involve any transference of energy to or from the void itself. If this is correct, though, then (ii) is false.

## VII. IS INTERACTIONIST DUALISM COMPATIBLE WITH (III), (IV), AND COE?

Before concluding the present discussion, a few words are in order regarding a certain proposal of Broad's (1925: 109), which warrants mention inasmuch

<sup>14</sup> Hence, Fair's (1979: 237) remark that 'the theory that causation is a matter of energy flow might be compatible with certain forms of dualist interactionism'.

<sup>15</sup> Broad (1925: 107–8), however, provides an apparent counterexample to transference theories involving a causal interaction between two purely physical entities.

<sup>16</sup> Fair (1979: 245–8) attempts to expand his transference account to cover cases of causation by omission, but to do so, he requires the use of counterfactuals that take him beyond a purely transference-based theory of causation.

<sup>17</sup> Not everyone is convinced that such cases do constitute genuine instances of causation. See e.g., Beebe (2004).

<sup>18</sup> The example is from Lewis (2004).

as it may seem to give dualists a way of rendering their position compatible not only with CoE, but also with (iii)—the claim that nothing non-physical has energy—and (iv)—the claim that the physical realm is a closed system. Having this option at their disposal would of course prove useful for dualists if the arguments offered against (iii) and (iv) above do not convince. The basic outlines of Broad's proposal are as follows: assuming both that the physical realm is a closed system in which energy is conserved and that no energy is transferred between the mind and any physical thing, Broad maintains that the mind could still causally influence bodily motion by 'determin[ing] that at a given moment so much energy shall change from the chemical form to the form of bodily movement. . . without altering the total amount of energy in the physical world'.<sup>19</sup> Since the kind of situation Broad envisions seems consistent with (iii), (iv), and CoE, it appears that further assumptions are needed to derive a valid argument against interactionist dualism from these premises. If the forgoing discussion is any indication, we might expect it to be rather difficult to find a set of assumptions that are strong enough to rule out Broad's hypothesis when conjoined with (iii), (iv), and CoE without also being at least as questionable as (iii) and (iv) were shown above to be. Thus, while Broad's model of psychophysical causation could be excluded by adopting a transference theory of causation such as that expressed in (ii), such theories are, as previously noted, open to a number of criticisms that would have to be answered before they could be reasonably accepted as grounds for rejecting Broad's proposal. One might naturally wonder whether the same might not be true of any other assumption added to (iii), (iv), and CoE with this end in mind.<sup>20</sup>

There is, however, at least one potential criticism of Broad's model of psychophysical causation that does not seem open to this objection, which is that while consistent with CoE, Broad's hypothesis still stands in at least apparent conflict with the law of conservation of momentum.<sup>21</sup> This is because it is difficult to see how the mind could alter the distribution of energy in a physical system, or the rate or time at which it is transferred or converted from one body or form to another, without also changing the speed or direction

<sup>19</sup> A similar suggestion is made by Larmer (1986: 281–2).

<sup>20</sup> To this point, Gibb (2010: 376, 382) offers up two premises that *would* yield a valid argument against interactionist dualism if conjoined with (iii), (iv), and CoE, but finds these premises 'dubious', primarily because they 'cannot be established by appealing to the energy transference theory of causation' and 'if these premises are instead to be inferred from facts within physics then it is unclear what these facts are'.

<sup>21</sup> Moreover, as an anonymous referee pointed out, even if Broad's proposal is consistent with all conservation laws, there may be other physical laws that it is not so easily reconciled with. Dualists who adopt Broad's proposal will hence have to either show that the mind can determine transformations of energy without violating any such laws or else argue that any laws that are thereby violated are not so fundamental or well established that it would be outlandish to think that they might need to be revised or replaced.

of the motion of certain bodies within the system. But if the consequent changes in the momentum of these bodies is indeed caused by some non-physical mental state (and thus cannot be fully accounted for in terms of the interaction of these bodies with other bodies), then it seems inevitable that those bodies whose motion is altered by mental causes will undergo changes of momentum that are not accompanied by the kind of compensatory changes in the momentum of other bodies needed to ensure that momentum is conserved.

Averill and Keating (1981) respond to this objection by arguing that since the law of conservation of momentum only requires that momentum be conserved in systems that are subject to no external force, Broad's proposal does not violate that law, for even if his hypothesis does imply that momentum is not conserved in the brain, or in the physical realm as a whole, the brain and hence the physical realm as a whole are, on Broad's view, subject to certain external, non-physical forces<sup>22</sup>, and so momentum needn't be conserved in those systems in order for the law of conservation of momentum to hold valid. While Averill and Keating provide no explicit argument for their claim that physical systems are, on Broad's view, acted upon by external, mental forces, their reasoning seems to be that since acceleration is by definition directly proportional to force, anything that causes an acceleration must do so by exerting some force on the accelerated object.<sup>23</sup> Hence, insofar as Broad's account entails that mental events cause accelerations in certain bodies, it must also require that the physical realm is subject to certain non-physical, mental forces whereby mental events produce such accelerations.

While this reasoning shows that Broad's model of psychophysical causation is consistent with both the law of conservation of momentum and CoE, it also shows that upon closer examination, Broad's model turns out to be inconsistent with both (iii)—the claim that nothing non-physical has energy—and (iv)—the claim that the physical realm is a closed system. For to allow that changes in bodily motion have non-physical causes is to allow that the physical realm is subject to non-physical forces, in which case (iv) is false. And since, as noted in Section II above, it is difficult to see how anything could exert a force without possessing energy, to allow that non-physical, mental events exert forces is (*pace* Averill and Keating) to allow that such events have energy, in which case (iii) is false. If this is correct, then contrary to first appearances, Broad's proposal cannot be adopted without rejecting both (iii) and (iv). While this does not mean that Broad's proposal is false, it does mean that interactionist dualists will have to look elsewhere if they wish to find a formulation of their position

<sup>22</sup> See also Larmer (1986: 282). Given the crucial role that this premise plays in Averill and Keating's argument, it is somewhat puzzling why Gibb (2010: 379) should cite them as providing a forceful defence of Broad's proposal against the objection noted above, while also claiming that Broad's proposal entails that 'an entity can cause... redistribution [of energy and momentum] without exerting a force'.

<sup>23</sup> Or, in the case of accelerations caused by omission, by failing to exert such a force.

that enables them to accept the conservation laws of energy and momentum along with (iii)—the claim that nothing non-physical has energy—and (iv)—the claim that the physical realm is a closed system.

## VIII. CONCLUSION

In light of the various objections we have now considered to the arguments against interactionist dualism from CoE, (ii), (iii), and (iv), the following conclusions seem warranted:

- CoE does not provide clear support for (1)—the claim that every physical effect has a sufficient physical cause. For there are formulations of interactionist dualism that reject (1) without violating CoE.
- Any direct argument against interactionist dualism from CoE will likely end up being invalid unless it makes use of certain additional premises (viz. (ii), (iii), and/or (iv)) that are disputable, and are not clearly supported by current science.

If these conclusions are correct, then interactionist dualism does not necessarily violate conservation laws, and it may likewise be possible for dualists to reject (1) without thereby setting themselves at odds with current science. Should the various attempts to respond to the Exclusion Argument by rejecting (2) prove inadequate, interactionist dualists can hence take comfort in the fact that an alternative response to the Argument may be available to them that is at least as consistent with current science as any of the assumptions on which the Argument itself is based.<sup>24</sup>

## REFERENCES

- Averill, E. and Keating, B. (1981) ‘Does Interactionism Violate a Law of Classical Physics?’, *Mind*, 90: 102–7.
- Battersby, S. (2013) ‘Chasing Shadows’, *New Scientist*, 218: 32–5.
- Beebe, H. (2004) ‘Causing and Nothingness’, in J. Collins, N. Hall and L. A. Paul (eds) *Causation and Counterfactuals*, 291–308. Cambridge: MIT Press.
- Broad, C. D. (1925) *The Mind and Its Place in Nature*. Paterson, NJ: Littlefield, Adams & Co.
- Crane, T. (2001) *Elements of Mind*. Oxford: OUP.
- and Mellor, D. H. (1990) ‘There Is No Question of Physicalism’, *Mind*, 99: 185–206.
- Denett, D. (1991) *Consciousness Explained*. Boston: Little, Brown & Co.
- Dobrescu, B. and Lincoln, D. (2015) ‘Mystery of the Hidden Cosmos’, *Scientific American*, 313: 32–9.
- Dowe, P. (2000) *Physical Causation*. Cambridge: CUP.
- Fair, D. (1979) ‘Causation and the Flow of Energy’, *Erkenntnis*, 14: 219–50.
- Feng, J. and Trodden, M. (2014) ‘Dark Worlds’, *Scientific American*, 23: 44–51.

<sup>24</sup> Many thanks to two anonymous referees for their helpful comments.

- Fodor, J. (1981) 'The Mind-Body Problem', *Scientific American*, 244: 114–23.
- Garcia, R. (2014) 'Closing in on Causal Closure', *Journal of Consciousness Studies*, 21: 96–109.
- Gibb, S. (2010) 'Closure Principles and the Laws of Conservation of Energy and Momentum', *Dialectica*, 64: 363–84.
- Hart, W. D. (1988) *The Engines of the Soul*. Cambridge: CUP.
- Kim, J. (1989a) 'Mechanism, Purpose, and Explanatory Exclusion', *Philosophical Perspectives*, 3: 77–108.
- (1989b) 'The Myth of Non-Reductive Materialism', *Proceedings and Addresses of the American Philosophical Association*, 63: 31–47.
- (1993) 'The Nonreductivist's Trouble with Mental Causation', in J. Heil and A. Mele (eds) *Mental Causation*, 189–210. Oxford: OUP.
- (1998) *Mind in a Physical World*. Cambridge: MIT Press.
- Koksvik, O. (2007a) 'In Defence of Interactionism', *Masters Thesis*, Monash University.
- (2007b) 'Conservation of Energy Is Relevant to Physicalism', *Dialectica*, 61: 573–82.
- Larmer, R. (1986) 'Mind-Body Interactionism and the Conservation of Energy', *International Philosophical Quarterly*, 26: 277–85.
- Leibniz, G. W. (c.1691/1896) 'Essay on Dynamics on the Laws of Motion', in A. Langley (trans.) *New Essays Concerning Human Understanding Together with an Appendix Consisting of Some of His Shorter Pieces*, 657–70. London: The Macmillan Company.
- (1714/1898) 'Monadology', in R. Latta (trans.) *The Monadology and Other Philosophical Writings*, 215–71. Oxford: The Clarendon Press.
- Lewis, D. (2004) 'Void and Object', in J. Collins, N. Hall and L. A. Paul (eds) *Causation and Counterfactuals*, 277–90. Cambridge: MIT Press.
- Montero, B. (2006) 'What Does the Conservation of Energy Have to Do with Physicalism?', *Dialectica*, 60: 383–96.
- Papineau, D. (2001) 'The Rise of Physicalism', in C. Gillett and B. Loewer (eds) *Physicalism and Its Discontents*, 3–36. Cambridge: CUP.
- Popper, K. and Eccles, J. (1977) *The Self and Its Brain*. New York: Routledge.
- Putnam, H. (1999) *The Threefold Cord: Mind, Body and World*. New York: Columbia University Press.
- Reich, E. (2010) 'Chameleon Cosmos', *New Scientist*, 207: 30–3.
- Salmon, W. (1997) 'Causality and Explanation: A Reply to Two Critiques', *Philosophy of Science*, 64: 461–77.
- Searle, J. (2004) *Mind: A Brief Introduction*. Oxford: OUP.

*Temple University, USA*