Conservation of Energy Is Relevant to Physicalism

Ole KOKSVIK†

ABSTRACT

I argue against Barbara Montero's claim that Conservation of Energy (CoE) has nothing to do with physicalism. I reject her reconstruction of the argument for physicalism from CoE, and offer instead an alternative reconstruction that better captures the intuitions of those who believe that there is a conflict between interactionist dualism and CoE.

INTRODUCTION

In "What Does the Conservation of Energy Have to Do with Physicalism?" Barbara Montero observes that many philosophers "hold that the conservation of energy law is inconsistent with interactive dualism", an inconsistency they take to "lead to an argument for physicalism" (2006, 385). Montero argues, however, that no valid argument against interactionism can be constructed without also taking on board, as extra premises, contentious claims about causation and the nature of entities that have energy.[1] Although the argument that results is valid, Conservation of Energy (CoE) becomes a redundant premise in it, and the title question of her paper must be answered with "nothing whatsoever" (395).

I agree with Montero that arguments for physicalism (or against interactionism) anchored in CoE are seldom carefully enunciated, and that reconstructing and evaluating these arguments is important. Although I must leave the discussion of that point for another occasion, I also take the argument from CoE to the falsity of interactionism to be inconclusive.

† School of Philosophy and Bioethics, Monash University and the Philosophy Program, Research School of Social Sciences, the Australian National University, Canberra ACT 2600, Australia. Email: ole.koksvik@anu.edu.au.

[1] I will often speak as if Montero considers an argument against interactionism, even though she mostly frames the discussion in terms of an argument in favour of physicalism. This inaccuracy is inconsequential since—on the premise that there are causal relations between the mental and the physical, a premise which figures in all the arguments Montero considers—an argument for physicalism is also an argument against interactionism.

Nevertheless, it seems to me that Montero's case for the outright *irrelevance* of CoE to interactionism does not succeed. Her argument does not adequately account for the intuitions that motivate the incompatibility claim, and it is therefore an unsatisfactory response to it.

In what follows, I briefly summarise Montero's argument and discuss the role of causation and closure in her argument. I propose an alternative interpretation of the notion of closure and present a plausible and valid argument against interactionism in which CoE figures non-redundantly.

PURPORTED IRRELEVANCE

Montero spends some time justifying her reconstruction of the argument for physicalism from CoE. Reconstructed in the preferred "general form" (388), the "Argument from the Conservation of Energy" (ACE) appears as follows:

1. Energy is conserved in any closed system.
2. The universe is a closed system.
3. There are causal relations between the mental and the physical.
4. Causation involves the transference of energy.

Thus: The mental is physical (385).

To make the argument conclusive, Montero points out that another premise has to be added: "anything with energy is physical" (393). When this last premise is appended, however, the first two premises are no longer needed to establish the conclusion. With the first two premises subtracted, and the remaining premises appropriately renumbered, we get what Montero calls the "Real Argument for Physicalism" (RAP):

1. There are causal relations between the mental and the physical.
2. Causation involves the transference of energy.
3. Anything with energy is physical.

Thus: The mental is physical (394).

This argument, which according to Montero "lies behind the supposed argument for physicalism from the conservation of energy", is valid, but CoE

is, of course, conspicuously absent from it (394). So, Montero concludes, CoE turns out to be "irrelevant to the question of whether the mental is physical" (395).

CAUSATION

Call a person who claims that interactionism is incompatible with CoE, and therefore false, an *incompatibilist*. It is plausible that the incompatibilist reasons in something like the following manner: "The physical world conserves energy. Therefore, if the dualist is right and the mind is outside the physical world, the mind is outside a system that conserves energy. If the mind then interacts with the brain, its interaction will violate conservation. But we know that conservation is not violated. So the mind cannot be outside the physical world and still interact with it."

As Montero points out, incompatibilist arguments are hard to find in explicit form. When considering the incompatibilist case, we are therefore left to reconstruct an argument on the incompatibilist's behalf. An obvious desideratum of such a reconstruction is that it should accommodate the intuitions the reasoning relies on, by giving them a central part to play in the argument against interactionism.

According to Montero, this cannot be done. She claims that there is *no* plausible and valid argument for physicalism in which CoE figures non-redundantly, and also that the argument which is really doing the work depends *inter alia* on the (contentious) claim that causation involves transfer of energy.

I reject both these claims. By way of backing up my rejection of the former claim, I provide below a plausible and valid argument against interactionism in which CoE figures non-redundantly, and which accounts for the incompatibilist's intuitions. But first I wish to motivate my rejection of the latter claim.

It is very implausible that the argument for physicalism from CoE should depend on the transference theory of causation, because it seems to go through perfectly well on a counterfactual theory. To make this vivid, consider the following thought experiment:

Imagine a non-physical omnipotent being. If such a being exists, it can bring something about in any logically possible manner whatever. The way it chooses to bring something about might be so alien to us that we should be forced to characterise the event as miraculous or, perhaps, magical. Suppose that the being uses its powers to move a single particle in a brain from one location to another. Then, the incompatibilist will argue, it is very likely that the energy-level of the brain after this occurred would be different to from what it was before. Then CoE has been violated.

This highlights the point that it is the end result—the change in the brain—that matters, and not the mechanism by which that change is brought about. The crucial ability attributed to the omnipotent being is the ability to cause a particle to move without this necessitating that the changes we usually observe in adjacent physical systems (changes in energy that are, in sum, equal, but of opposite sign) occur as well. Interactionism holds that the mind has that ability.[2]

Interactionism claims that changes in the brain depend counterfactually on changes in the mind. Suppose that interactionism is true, and that because of the mind a particle in the brain moves when it otherwise would not have. Then, the incompatibilist will argue, it is very likely that the mind has changed the energy-level of the brain, and that CoE has therefore been violated. But CoE is never violated. So changes in the brain are very unlikely to depend counterfactually on the mind, and interactionism is very likely false. That this line of argument is open to the incompatibilist shows that the counterfactual theory of causation is quite enough to get an argument from CoE against

interactionism started.[3]

CLOSURE

Before constructing a plausible and valid argument for physicalism in which CoE figures non-redundantly, we must consider how a crucial notion should be understood. CoE says that in any *closed* system, energy is *conserved*. The latter of the two italicised terms is not too hard to define; energy is *conserved* in a system if the total quantity of energy in the system is unchanged over time, if "there is at every moment in time the same total amount of energy" (Dieks 1986, 90). But what is it for a system to be *closed*?

Montero's interpretation of this notion plays a crucial role in her argument for the claim that there is no plausible and valid argument for physicalism in which CoE figures non-redundantly. She writes:

> [T]he notion of being a closed system here is distinct from the notion of being causally closed as it is used in the [causal] argument for physicalism ... [where] the physical world is said to be causally closed, roughly, if every physical effect that has a sufficient cause has a sufficient physical cause. Here, 'a closed system' means that the system neither affects nor is affected by anything from outside of it (385, n. 12).

An important part of the intuition driving the incompatibilist is almost certainly that if the dualist is right, the mind lies outside a conservative system. Montero contests this. Her argument proceeds in two steps.

In the first step, Montero argues that the incompatibilist cannot suppose that the *physical world* is closed in the sense required by CoE.

---

[3] Not only is the transference theory of causation unnecessary to get the argument off the ground, but its inclusion seems to make the argument prove too much. As Montero acknowledges, an argument against interactionism in which the transference theory of causation forms part is also an argument against epiphenomenalism. That conclusion runs contrary to the received view of what epiphenomenalism entails. As the quotations in Montero's paper illustrate (384), proponents of the incompatibility claim do not worry that energy would "disappear" if epiphenomenalism were true, and epiphenomenalists typically do not take the brain to have to expend energy to give rise to the mind. Campbell, for example, says in regards to "the production of spiritual effects by material causes" that "it is no part of the conservation principle that the production of non-material effects requires physical energy" (1970/1980, 53).

Given her notion of closure the reason is obvious: claiming that the physical world were closed in her sense would be to beg the question outright against interactionist dualism. If the physical world is not affected by anything outside of it, and if the mind is not physical, it follows that the mind does not affect anything in the physical world. Then interactionism is false. To avoid begging the question, Montero argues that the only system the incompatibilist can suppose is closed is the universe as a whole: the system of all there is, physical or non-physical (386).

Which systems are closed in this sense and which conserve energy are two separate questions. Perhaps the incompatibilist could argue that though the only *closed* system is the whole universe, there is a conservative *sub-system* to be found. In particular, could the incompatibilist claim that the *physical* sub-system is conservative? If the mind was thought to be inside the *closed* system but *outside* the conservative one, an argument against interactionism could still get off the ground. In the second step Montero argues that the empirical evidence can only legitimately be taken as evidence for the claim that energy is conserved in the *whole* of the closed system and *not* as evidence for the claim that energy is conserved in the physical sub-system.[4] That seals the case. If the mind is not only inside the closed system but inside the conservative system as well, no valid argument for physicalism can be advanced on the basis of CoE without denying that the mind has energy of its own. If it does, its participation in causal transactions could still conserve energy. This leads to the conclusion that the real argument for physicalism has as its premises the claims that mind and body interact causally, that causation involves transference of energy, and that anything with energy is physical.

[4] I find her argument unconvincing (more on this later).

Since, as we have just seen, Montero's notion of closure plays an integral role in the argument for the claim that there is no plausible argument for physicalism in which CoE plays a non-redundant part, it is reasonable to look for an alternative interpretation of the notion. But what should it be?

For the purposes of CoE, the standard interpretation of closure has it that a closed system is such that energy *does not flow* or *is not transferred* into or out of it (Dowe 2000, 95; Halliday *et al.* 1997, 167). Call such a system an *e–closed* system. One indication that e–closure is the right notion is that it yields a result that accords well with the "slogan version" of the law. The slogan version says that energy is neither created nor destroyed, or that it "cannot magically appear or disappear" (Halliday *et al.* 1997, 168). If a system is e–closed—if it has energy-tight borders—and if energy is neither created nor destroyed within it, then the total quantity of energy within the system will remain unchanged, which is what CoE claims. So e–closure is a good candidate notion.

There is, however, a complication.[5] Imagine an experiment examining a purportedly e–closed system revealing that energy has been transferred to an entity in the system from an entity outside the system. The correct conclusion to draw from this would be that the system was not e–closed after all. On the other hand, the discovery that the energy level of an entity inside the system *just rose* would rightly be taken as a case of an e–closed system where energy was *not conserved*, and would therefore constitute a violation of CoE. The distinction between the two cases seems to rely on energy being identifiable and re-identifiable through time. If there is no way to determine that *this* 'bit' of energy is the same as *that*, the two outcomes would seem to be indistinguishable from one another: we would be unable to distinguish a situation where the law's application condition failed to obtain from a counter-

[5] I am grateful to Toby Handfield for bringing this to my attention.

instance to the law.[6]

However, the suggestion that energy has identity over time appears to be in dissonance with current physical theory (Dowe 2000, 55-59; Dieks 1986, 87-91). Dowe, for example, concludes that "energy generally doesn't have identity through time" (59), and Dieks states that "the idea that energy ... possess[es] an identity of [its] own which is retained throughout physical interactions is already foreign to classical physics", and in quantum mechanics "[t]he mere *ascription* of an identity to the energy elements ... entails consequences which are observably wrong (88, 89). It therefore seems that we may be forced either to relinquish the use of e–closure in the formulation of CoE, or else to conclude that CoE is immune, in a particular way, to counterexamples, raising questions about its substantiveness.

Despite this, I think there are good reasons to retain the notion of e–closure and the formulation of CoE of which it forms part. First, even if this particular formulation of e–closure must be relinquished, it is not clear that another cannot be put in its place that retains its use in CoE. Dieks has suggested a mathematically defined *current* of energy that does not rely on energy being identified. With this notion e–closure can be restated in terms of the absence of incoming or outgoing currents.[7] Second, e–closure enables the formation of an argument which accounts for the incompatibilist's intuitions and preserves the relevance of CoE. So there is good reason to think that,

---

[6] It has been suggested to me (by an anonymous referee) that if the description of the latter case were amended to include that a closely neighbouring system lost roughly as much energy as the purportedly closed system had gained, we would be justified in inferring that it was not closed after all. I have worries about this suggestion that I see no reason to discuss here; it suffices to note that if this is correct the notion of e-closure does not depend on identification and re-identification of energy, and so the argument I offer below is not threatened by the complication under consideration.

[7] "We can consider energy and momentum densities as functions of space and time coordinates and we can calculate the changes of the densities if we vary the values of these coordinates. This gives us a rate of change in every direction. In this way we can mathematically define a 'current' of energy and momentum. In view of the conservation laws the total 'current' for an isolated system must vanish" (90).

at least when reconstructing an argument on the incompatibilist's behalf, e–closure, or a notion very much like it, should be employed. In what follows I assume that a closed system can be defined either in terms of flow or transfer of energy, or in terms of some closely related notion that does the same job.

RELEVANCE REASSERTED

I claimed above that Montero's strategy fails to adequately account for the intuitions that are likely to be central in motivating the incompatibilist. One of these is that, according to the dualist, *the mind lies outside a conservative system*. Another is that if it does, *the mind's interaction with the conservative system would violate conservation.* Once the notion of closure in CoE is understood as e-closure it becomes possible to construct an argument in which both these intuitions play a role:

1* Energy is conserved in any e–closed system (assumption).
2* The physical world is an e–closed system (assumption).
3* If a non-physical mind changes a physical system, it changes its energy-level (assumption).
4* If the energy-level of a physical system is changed by a non-physical system**,** energy is not conserved in the physical world (assumption).
5* Therefore: It is not the case that a non-physical mind changes a physical system. [8]

1* and 2* take care of the intuition that if dualism is true, the mind is outside a conservative system. Dualism holds that the mind is not a part of the physical world, and from 1* and 2* it follows that the physical world is conservative. In turn, 3* and 4* jointly express the incompatibilist's second central intuition, that the interaction of a non-physical mind with a conservative system would violate conservation.[9]

---

[8] Many thanks to David Chalmers and Toby Handfield for very helpful suggestions on this argument.

[9] That is, they do on the assumption that the changes mentioned in the antecedent of 4* are sufficient to cause downstream behavioural consequences, an assumption left implicit here.

Note that 1* is just CoE, on what I have argued is the best way to understand a closed system. 2* is a substantive assumption, but it is one we can expect the incompatibilist to be happy to make, and unlike the claim that the physical world is closed in Montero's sense, 2* does not beg the question against interactionism.

Montero uses a distinction between two interpretations of CoE to argue that the claim that energy is conserved in the physical world is not supported by our evidence. The "principle of physical conservation" (386) holds that energy is conserved among the physical components of the closed system, whereas the unrestricted version makes no such qualification. She goes on to claim that "while physics gives us reason to believe the [unrestricted version], it does not seem to give us reason to believe [physical conservation]" (386).

By way of backing this up, Montero argues that restricted version cannot be accepted "without question" while the unrestricted version can, because the former is "a philosophical principle rather than a law of physics" (387-88). It seems to me, however, that any evidence we have for CoE is evidence for the restricted version, but not for the unrestricted one, and that CoE regards the physical components of a system.

Montero considers this possibility: "One might argue that as all close observations that support the conservation of energy have been of entirely physical systems, all of our evidence for the conservation of energy is *also* evidence for physical conservation" (388, emphasis mine). She goes on to object that "when evidence for a theory is consistent with another theory ... [it does not follow that] the other theory is justified by this evidence" (388).

This objection seems to get things the wrong way around. The question is not whether the evidence is *also* evidence for the restricted version of CoE; given that the experimental evidence in favour of CoE has resulted from observations of entirely physical systems, it is the generalisation to the

unrestricted version of CoE that is unwarranted by our observations. The evidence we have for CoE is *only* evidence for the restricted version.

It is perfectly true, of course, as Montero points out, that an experiment with mind-body interaction which showed that energy was not conserved among the physical components of the system would provide a counter-example to the restricted version of CoE but not to the unrestricted version, since one could then stipulate that energy was conserved in the entire system, but not in the physical subsystem. But that does not show that the unrestricted version of CoE is what is supported by our present evidence; it just shows that CoE is not as well established as some might have thought that it is, since it has not been tested in the full range of circumstances. At least insofar as CoE is an empirical, *a posteriori* discovery, it is shown to be a thesis about the physical part of a system by the evidence adduced in its favour.[10]

CONCLUDING REMARKS

The reconstructed argument is not an argument for physicalism, since epiphenomenalist dualism is compatible with its conclusion. It *is* an argument against interactionist dualism. But it is not a knock-down argument. An obvious way to reject it is to claim that the mind causes changes in the brain exclusively by *reconfiguring* it in such a way that its level of energy is *not* changed.[11] If that is true, 3* is false and the argument fails.[12]

Can a non-knock-down argument against interactionism show that CoE is relevant to physicalism? It can. If the four premises are accepted, the reconstructed argument can be made into an argument for physicalism by the addition of a premise to the effect that the mind changes the body and the

---

[10] I am grateful to Jessica Wilson for convincing me that this sentence should be conditional.

[11] See e.g. (Campbell 1970/1980, 52-53; Broad 1925, 103-09).

[12] There are also promising strategies for resisting the argument that depend on rejecting the final premise, but since it is not here my concern to consider how an interactionist should respond to the argument, discussing these options would take me too far afield.

body changes the mind, a premise accepted by all but a few philosophers. And even if 3* is rejected in the way just outlined, CoE is still relevant to physicalism. The vast majority of changes that can be made to a physical system will change its energy-level, so the set of reconfigurations that do *not* change the energy-level of the brain is small compared to the set of changes that do. Rejecting 3* therefore places a significant constraint on the interactionist theory. Since interactionist dualism is one of its most important competitors, this ensures CoE's relevance to physicalism.

The argument I have presented is valid, it accounts for the incompatibilist's central intuitions, and CoE forms an integral part of it. Neither the transference theory of causation nor the claim that anything with energy is physical figure in it. Those two claims are both more general than is required. The question of whether the mind has energy, and the question of whether causation between two physical systems or events always involves transference of energy are both beside the point. What matters for the purposes of an argument from CoE is just that mental–physical causation is not characterised by energy-transfer. The premise which states that the physical world is an e–closed system ensures that that is true.[13]

For all these reasons, the argument presented above is a much more reasonable reconstruction on behalf of the incompatibilist than is RAP. It does not show that the truth of CoE ultimately proves interactionism false. But it does show that CoE is relevant both to interactionist dualism and to physicalism.

---

[13] Another way to resist the argument is, of course, to deny that the physical world is e–closed, by supposing that the mind has energy of its own. But just as before, this restricts the interactionist theory significantly, and therefore certainly ensures the relevance of CoE.

* Many thanks to John Bigelow, David Chalmers, Ian Gold, Toby Handfield, Graham Oppy, Jessica Wilson and anonymous *dialectica* referees for many helpful comments on earlier versions of this paper. Some parts of this paper were included in my talk at the 2007 meeting of the Australasian Association of Philosophy in Armidale, and I am grateful to the audience there for comments. Finally, I gratefully acknowledge the support I received from the Faculty of Arts at Monash University in the form of a Postgraduate Publication Award.

REFERENCES:

Broad, C. D. 1925. *The Mind and its Place in Nature*. London: Routledge & Kegan Paul.

Campbell, K. 1970/1980. *Body and Mind*. Notre Dame: University of Notre Dame Press.

Dieks, D. 1986. Physics and the Direction of Causation. *Erkenntnis* 25 (1):85-110.

Dowe, P. 2000. *Phyiscal Caustion*. Cambridge: Cambridge University Press.

Halliday, D., Resnick, R., and Walker, J. 1997. *Fundamentals of Physics: Extended*. Fifth ed. New York: John Wiley & Sons.

Montero, B. 2006. What does the Conservation of Energy Have to Do with Physicalism? *dialectica* 60 (4):383-96.