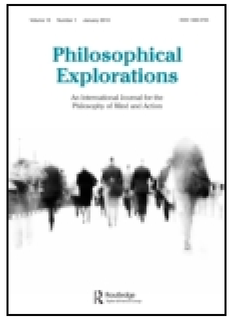

No Perfect Pass: How The Energy Conservation Objection Against Dualism Turns Out To Be Physicalism's Own Goal

Journal:	*Philosophical Explorations*
Manuscript ID	RPEX-2018-0080
Manuscript Type:	Essay Prize Submission
Keywords:	Dualism, Physicalism, Energy Conservation, Neuroscience

No Perfect Pass: How The Energy Conservation Objection Against Dualism Turns Out To Be Physicalism's Own Goal

Abstract

In this paper, I aim to show (1) that the principle of energy conservation (PEC) cannot be used as an *a priori* argument against dualism; (2) that PEC constitutes a problem for physicalism because energy is probably not conserved in brains; and (3) to show that even if energy is conserved in brains, dualism is still the better framework to account for human volitional actions. To do that, I will first formulate a proper *a priori* version of the widely shared 'objection from energy conservation' (OEC). Second, I will refute its central premise, namely that energy is necessarily conserved. I then proceed to "turn the tables" on physicalism, i.e. I seek to show that *empirically,* energy is *probably not* conserved in brains, and explore the resources of physicalism to deal with such a scenario, which turn out to be scarce.

1. The EC objection: a proper formulation

It is widely believed that the principle of energy conservation[1] (PEC) poses at least serious difficulties to dualism[2]. Interestingly, though, it is difficult to get a proper formulation of that 'objection from energy conservation' (OEC) in the literature. The main problem is that worries about energy conservation are being confused or conflated with some version of the 'causal nexus problem'[3]. A much cited passage in Daniel Dennett's *Consciousness Explained* follows exactly that pattern (Dennett 1991, 34-35)[4]:

> [T]he return signals, the directives from mind to brain (…) are not physical; they are not light waves or sound waves or cosmic rays or streams of subatomic particles. *No physical energy or mass is associated with them. How, then, do they get to make a difference to what happens in the brain cells* they must affect, if the mind is to have any influence over the

[1] Though less prominent, momentum conservation is, as a physical principle, on equal footing with energy conservation. In fact, the earliest science-based objection against substance dualism appealed to momentum conservation (e.g. Garber 1983; Leibniz 1985). Since then, it has mostly been overlooked, and, as will be seen, thwarts many naïve attempts to make dualism physics-friendly.
[2] By dualism, I mean the thesis that there are mental entities which exist on their own (i.e. not grounded in, supervenient on or emergent from physical things). I avoid the term 'substance dualism' because, while I tend to call the mind/soul a substance, I am skeptical about calling the body a substance.
[3] The causal nexus problem consists in the intuition that there does not seem to be any 'causal interface' between non-physical and physical entities that would allow the non-physical entities to interact with the physical world.
[4] For other examples see e.g. Mcginn 2000, 92; Westphal 2016, 41-44.

> body? *A fundamental principle of physics is that any change in the trajectory of any physical entity is an acceleration requiring the expenditure of energy*, and *where is this energy* [in mind-brain-interaction] *to come from*? It is this principle of the conservation of energy that accounts for the physical impossibility of "perpetual motion machines", and the same principle is apparently violated by dualism. This confrontation between quite standard physics and dualism has been endlessly discussed since Descartes's own day, and is widely regarded as the inescapable and fatal flaw of dualism. [Italics and brackets added]

Dennett speaks of a "confrontation between quite standard physics and dualism", but upon closer examination, no such confrontation appears to be present. His question where the energy for mind-brain-interaction is to come from insinuates that it cannot come from the mind[5], because the mind lacks physical properties. But if the mind indeed lacks physical properties, and this lack leads to causal ineffectiveness in the physical world, then Dennett's objection could be reframed as "The mind cannot make a difference in the physical brain, because it lacks physical properties." Note that in this reformulation, PEC does not figure at all. Therefore, I take it that what Dennett delivers is not a version of OEC, but a version of the causal nexus problem. To see even more clearly that OEC and the causal nexus problem are distinct, consider that it would still be an open question whether mental interaction is at odds with energy conservation if the mind's causal efficacy were warranted.

A version of OEC that addresses the real issue more aptly can be found with John Searle:

> Physics says that the amount of matter/energy in the universe is constant, but substance dualism seems to imply that there is another kind of energy, mental energy or spiritual energy, that is not fixed by physics. So if substance dualism is true then it seems that one of the most fundamental laws of physics, the law of conservation, must be false. (Searle 2004, 42)

Searle's argument is interesting in two respects. First, his version of PEC ("the amount of matter/energy in the universe is constant") is probably the one most widely used by non-physicists (though differs substantially from the one modern physics holds, see below). Second, he aptly points out that the crux lies in an apparent contradiction between dualism and PEC, assuming that in this case dualism will have to yield, given the fundamentality of PEC. Formally, the argument could run as follows[6]:

---

[5] I here use the terms 'mind' and 'soul' interchangeably.

[6] After replacing 'mental energy' by '(physical) energy'. It seems that talk of energy requires ascribing pertinent mathematical properties to the energy-bearer. It is unclear how this can be done with mental energy (cf. Pitts 2018, 5). If it were possible, the question would be what then distinguishes mental energy from physical energy, apart from the different types of bearers. But even laying aside those issues, a rise in energy in the physical world will be

<u>OEC</u>:

P1 The amount of energy in the physical universe does not change. (PEC)

P2 Dualistic minds are purely non-physical[7] and hence not part of the physical universe.

P3 Brains are parts of the physical universe.

P4 If dualistic minds caused brain events, they would change the brain's energy content. (*ex hypothesi*)

P5 If dualistic minds caused brain events, they would change the amount of energy in the physical universe. (from P1-P4)

C Therefore, minds are not non-physical (from P1, P2, P5 by *modus tollens*)

There is, however, a problem with P1. Cosmologists are split over the question whether the total amount of energy in the universe can even be calculated (cf. Pitts 2004a, 2004b, 2009). Therefore, and for reasons of relevance to the philosophy of mind, in P1 ought to be modified as follows:

P1* Energy is conserved in brains.

(Talk of an unchanging energy amount does not make sense in the case of brains, because brains, being open system, very obviously change their energy amount constantly.)
But now we run into another problem. The claim that energy is conserved in brains without having ascertained it empirically can only rely on inductive inference from other (bio-)physical settings. At this point, a defender of OEC might claim that this inductive approach is a proven, genuinely scientific method: if energy conservation holds in the rest of nature, it most probably holds in brains as well, so why even consider the option of even occasional violations of PEC? This reasoning has a fatal flaw, for it overlooks that in the case of dualistic interactions, one deals with the hypothesis of a non-physical mind, which cannot be refuted by appealing to energy conservation on pain of begging the question against dualism. It might just be that the induction works wherever no mental interaction is present. So in this case, empirical data seem to be vital to support the anti-dualist's case, but contrary to the assertions of some, (e.g. Montero and Papineau 2016, 188), such data are not available with respect to the pertinent brain processes[8]

---

measured in *physical* energy. Thus, if Searle's 'mental energy' expended by the mind is to make any difference in the physical world, it needs to be converted into physical energy.

[7] I am very much aware that there are dualistic accounts which construe the mind as partly physical (e.g. Collins 2011).

[8] To be sure, Eccles (Popper and Eccles 1977) and Eccles & Beck (Beck and Eccles 1992, Eccles 1994) did some work on this question. Their investigations are, however, motivated by dualism; therefore, they are interested in finding 'spots' where the mind might interact rather than in canvassing and scrutinizing all possible physical sources for the brain events in question.

(see below). And as long as they are not, claiming that energy conservation applies to the brain without prior refutation of the existence of souls on independent grounds is to beg the question against dualism.

Thus, it seems that what the anti-dualist needs in order to get an *a priori* argument against dualism is the claim that energy is *necessarily* conserved in brains[9]. One might suspect that this is too strong a formulation and therefore vulnerable (which indeed it is, see next section). All the more it is surprising that dualist efforts have concentrated on P2 (claiming that souls are partly physical, see Collins 2011 or Hart 1994) or P5 (claiming that souls can interact with the brain without violating energy conservation, e.g. Ducasse 1960, Broad 1937, Lowe 1992, Dilley 2004, Meixner 2008, Gibb 2010, White 2016).

We are now in a position to formulate a modified (and shortened) version OEC:

OEC':

P1' Energy is necessarily conserved in brains (PEC')

P2' Dualism entails that energy is not conserved in brains.

C' Therefore, dualism is false (by *modus tollens*).

It is important to make clear what energy conservation means in this context. It does not mean, of course, that there are no energy changes in the brain; as a living organ, the brain is a swirling sea of metabolism and neural activity, and in constant matter and energy exchange with its immediate physiological environment. In other words, the brain is an open system. How can energy conservation be verified for an open system? I accept Robin Collins's (Collins 2008) BPEC (boundary principle of energy conservation) as a suitable candidate for open systems. According to BPEC, "the rate of change of energy (…) in a closed region of space is equal to the total rate of energy (…) flowing through the spatial boundaries of the region." (ibd., 34). Thus, energy will be conserved in a brain if the energy change in the brain can be accounted for by the amount of energy flowing through the brain's spatial boundaries. This amounts to there being physical entities outside the brain which confer the energy to or take it from the brain, assuming

[9] A similar reasoning can be found in (Larmer 2009, Larmer 2014; Plantinga 2007; all pertaining to divine interactions); see also (Von Wachter forthcoming) for a parallel and closely related reasoning with respect to the causal closure of the physical world. By the way, it seems that it is such a modal claim that many philosophers and scientists have in mind when they speak of a *law of nature*, just as energy conservation is of often referred to as a 'law'. To avoid that potentially misleading language, I chose the term 'principle' instead.

that only physical entities can carry energy. In summary, PEC', together with BPEC, claims that there necessarily are physical causes which explain the energy changes in an open system.

At this point a remark is in order. With the advent of quantum mechanics and the Heisenberg uncertainty in particular, a certain 'blur' seems to affect energy and momentum conservation at the submicroscopic level. The Heisenberg uncertainty says, roughly, that holding the momentum of a particle fixed, its position is 'blurry', i.e. uncertain within certain limits. This in principle allows for particle movements without the expenditure of energy, and the movement of one particle without energy expenditure might set in motion movements of other particles which do involve energy changes, maybe in the way a small stone may trigger an avalanche. If such processes happened in the brain, they would "mimick" a violation of PEC' (because it would seem as if there were no outside source for the energy increase), while in fact there is none. I take it that if such quantum events occured in the brain – which is a matter of debate (see Beck & Eccles 1992 vs. Wilson 1999) – this would count as energy conservation.

2. Energy conservation in modern physics

I will now argue against P1' by examining what modern physics has to say about energy conservation. Perhaps surprisingly, modern physics does not construe energy conservation in the way P1 does, namely *globally* (i.e. pertaining to the whole universe). One would search physics textbooks in vain for such a statement[10]. The main reason is that there might be no global conservation. It can be obtained if and only if the surface integral[11] of the whole universe becomes zero. This can, however, fail even without the interaction of non-physical entities; for example, if the universe were an infinite Euclidean space (which is one of three possible standard Big Bang cosmology models), then the amount of energy would be infinite, and energy changes would not make any difference (see e.g. Pitts 2004a, 2004b)[12].

[10] This does not seem to be the only case of a 'physics myth'. With respect to the closely related question of the causal completeness of the physical world, Sophie Gibb (2010, 366) writes: "[I]f Completeness is a working hypothesis of current physics, then it is one that is left wholly implicit – the principle is not referred to in any physics textbook." (See also Papineau 2000, 184-185).

[11] The surface integral is obtained as the volume integral of all energy fluxes $\frac{dJ_x}{dx} + \frac{dJ_y}{dy} + \frac{dJ_z}{dz}$.

[12] Assuming that the matter in the universe is more or less equally dense (and nonzero) everywhere, the adding up of the total amount gives infinite energy and the surface integral is not 0. Another way global energy conservation can fail is as follows: The universe might be a spatial 'quilt', i.e. it can only be described mathematically by dividing it into 'patches', each with its own mathematical description with respect to energy and momentum. At the 'seams' between the patches, there might be discontinuities to the effect that an integration of the 'patches' is impossible (Brian Pitts, personal conversation).

Noether theorem an Lagrangians

Modern physics derives energy conservation (just as the conservation of any physical quantity) in the first place from the first Noether theorem (Noether 1918; Goldstein 1980, ch. 12-7), which can be rendered informally as:

> If a system has a continuous symmetry property, then there are corresponding quantities whose values are conserved in time.

And its converse:

> If in a system there are quantities whose values are conserved in time, then there is in the system a corresponding continuous symmetry property.

Logically speaking, the Noether theorem (NT) entails a biconditionality between continuous symmetry[13] and conserved quantities:

Continuous symmetry ↔ conserved quantity(ies)

Thus, NT entails that if (and only if) a physical system does not have a continuous symmetry property, then there necessarily are quantities whose values are not conserved in time; conversely, if (and only if) some quantity fails to be conserved in time, there cannot have been a continuous symmetry property.

Another way of checking energy (and momentum) conservation is by using the Lagrangian[14] and the Euler-Lagrange equations[15]. Fortunately, though, we do not need to do those calculations here, because energy and momentum conservation can be read off the Lagrangian: Energy is conserved if the variable t (time) does not figure in the Lagrangian[16]; momentum is conserved if none of the variables x, y or z (place)[17] figure in the Lagrangian[18]. Consider the following simple

[13] Continuous symmetry - as opposed to non-continuous (= discrete) symmetry – can perhaps best be grasped by an example. A sphere has continuous symmetry, because applying a symmetry operation (e.g. case rotating it around an axis) yields symmetry upon choice of any values for the operation (in this case rotating the sphere by 360° results in as much symmetry as rotating by 180°, 112°, 97,35° and so on). A cube, however, has discrete symmetry: rotating it by n*90° yields symmetry (with n being an integer), but not rotation by, e.g., 112°.

[14] The Lagrangian is a functional that satisfies the principle of least action. In physics, it is used to reformulate classical mechanics problems with generalized coordinates.

[15] The Euler-Lagrange equations are second-order partial differential equations whose solutions are the functions for which a given functional is stationary. The Euler-Lagrange equations are used to calculate the solutions at which a given Lagrangian is 'stationary', i.e. the system's action is 'most' or 'least'.

[16] i.e. if the Lagrangian does not explicitly depend on time.

[17] There may be one, two or three spatial variables in the equation, depending on the n-dimensionality of the system in question.

[18] i.e. if the Lagrangian does not explicitly depend on place.

example[19]; it consists of a particle with mass m moving one-dimensionally (z-axis) in the gravitational field of the earth[20].

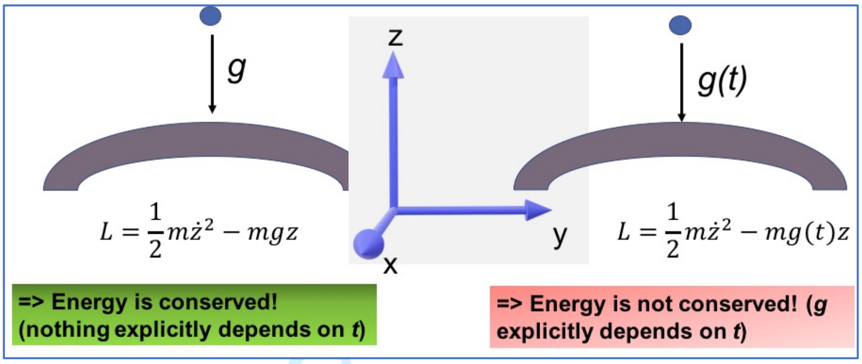

Figure 1: Particle in one-dimensional gravitational field (© Alin Christoph Cucu)

Since g is indeed a constant, i.e. time-independent, the Lagrangian of the system is

$$L = \frac{1}{2}m\dot{z}^2 - mgz$$

Nothing in this equation explicitly depends on t, so energy is conserved[21]. However, if g were a function of time, the Lagrangian would change into

$$L = \frac{1}{2}m\dot{z}^2 - mg(t)z$$

Now, something in the Lagrangian, namely g, explicitly depends on time; thus, as can be verified by calculation[22], energy is not conserved.

What is the connection between Noetherian symmetries and Lagrangians? We can see it by way of another simple example. Suppose two spheres with finite mass collide:

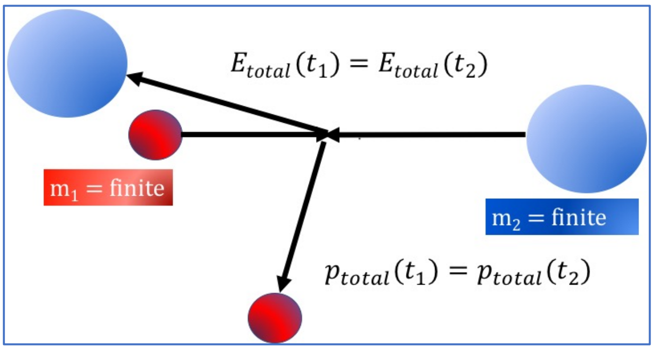

Figure 2: Collision of two spheres with finite mass (© Alin Christoph Cucu)

[19] cf. Pitts 2018

[20] Whose 'acceleration constant' is the familiar $g = 9.81\frac{m}{s^2}$.

[21] In fact, momentum is not conserved, because this Lagrangian explicitly depends on place (z); if one integrates the momenta of the earth and the gravitational field, it turns out to be conserved.

[22] cf. Pitts 2018

As can be calculated with simple classical equations, momentum[23] and energy[24] are conserved in this system. But we can also understand the conservation of momentum and energy by applying the Noether theorem:

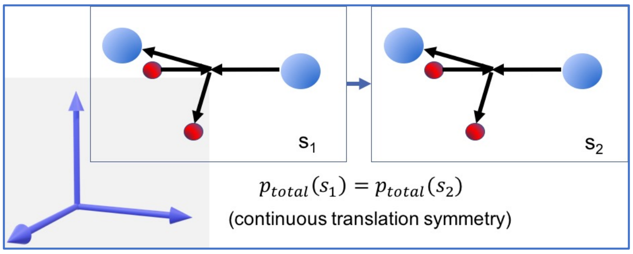

Figure 3: Space translation symmetry of the two-spheres system

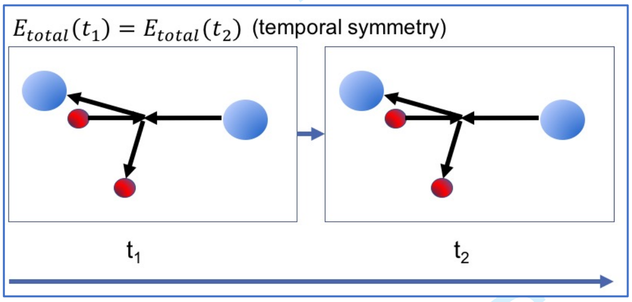

Figure 4: Time translation symmetry of the two-spheres system (© Alin Christoph Cucu)

Roughly speaking, since the system can be continuously moved[25] in space without any change in the Lagrangian description[26] of its elements, momentum is conserved (figure 3). Likewise, since the system can be continuously 'moved' in time without change, energy is conserved[27] (figure 4).

How the mind can make energy conservation fail: a simple model

So far, we have seen that energy and momentum are conserved in a system *on the condition* that the system's temporal and spatial symmetries hold. This, in principle, opens up the possibility for the symmetries to be broken and energy/momentum to not be conserved. One might wonder how this can happen. If a system's symmetries are broken (i.e. if the Lagrangian explicitly depends on time or place), then surely a bigger system can be defined whose symmetries hold? Recall the above example of the particle in the Earth's gravitational field. The seeming non-conservation of momentum can be fixed by considering the Lagrangian of the Earth-plus-particle system. So how can we understand a *genuine* violation of conservation principles? It seems that thos genuine violations will have to be caused by something non-physical. I will now present a simple 'toy model' for symmetry-breaking mental interactions with the brain. Consider a universe which

[23] $m_1\, \vec{v}_1(t_1) + m_2\, \vec{v}_2(t_1) = m_1\, \vec{v}_1(t_2) + m_2\, \vec{v}_2(t_2) => p_{total}(t_1) = p_{total}(t_2)$
[24] $\frac{1}{2}m_1\, {v_1}^2(t_1) + \frac{1}{2}m_2\, {v_2}^2(t_1) = \frac{1}{2}m_1\, {v_1}^2(t_2) + \frac{1}{2}m_2\, {v_2}^2(t_2) => E_{total}(t_1) = E_{total}(t_2)$
[25] Which is a symmetry operation.
[26] Of course, the spatial coordinates change; but the main advantage of a Lagrangian is precisely to give a coordinate-independent description of a physical system.
[27] In more technical terms, momentum conservation depends on space-translational invariance (= symmetry) and energy conservation depends on time-translational invariance.

consists only of an infinitely hard[28] sphere with finite mass and an infinitely hard wall with infinite mass. Sphere and wall move toward each other, so that they collide and that the sphere bounces back.

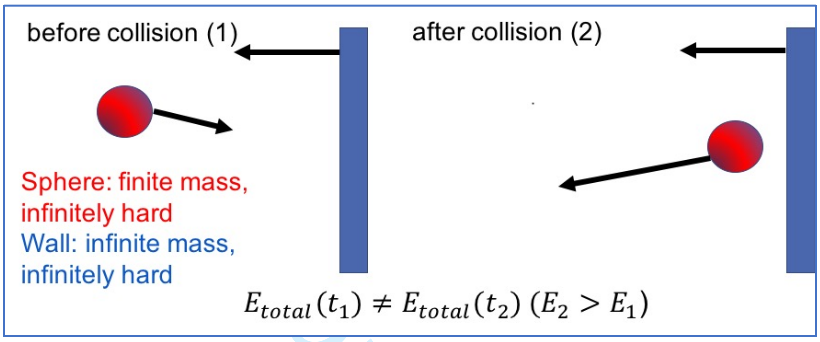

Figure 5: 'Toy model' of mental interaction with a massive object (© Alin Christoph Cucu)

In this scenario, the sphere's kinetic energy increases over time; the kinetic energy of the wall, however, is at all times *infinite* (due to its infinite mass[29]), which means that there cannot be a sensible Lagrangian description of the sphere-plus-wall-system. In other words, the wall is really *no part of the system*. Thus, the overall energy of the system increases in time. Put differently, the system (i.e. the sphere alone) is neither continuously symmetrical with respect to space nor with respect to time[30]; thus, in accordance with the Noether theorem, neither energy nor momentum are conserved. This simple model seems to be a close analogy for mental interactions, because just like the infinitely heavy wall, a purely non-physical mind cannot be described physically/mathematically in any sensible way. It should be construed as a non-physical 'out-of-system' agent which acts on the system (i.e., the brain), thereby breaking spatial and temporal symmetries and changing energy and momentum.

Let us sum up. Energy and momentum conservation can fail if a system's temporal and/or spatial symmetries are broken by an influence that cannot be described in physical terms. Whether such influences exist and whether they can exert causal influence are of course open questions (the latter being the causal nexus problem); by the same token, it begs the question against dualism to claim *a priori* that mental influences either do not exist or are causally ineffective. For the purpose of the present paper, it suffices to conclude that modern physics construes energy and momentum conservation in such a way that they in principle allow non-

---

[28] The infinite hardness for both objects is needed to disregard energy changes due to plastic deformation.

[29] One obtains non-relativistic kinetic energy by calculating $E_{kin} = \frac{1}{2}mv^2$. With infinite mass, one obtains $E_{kin} = \frac{1}{2}\infty v^2 = \infty$.

[30] Spatial asymmetry can be seen by moving the light sphere horizontally; its spatial relation to the out-of-system-wall (even if it stood still) then changes. As regards, temporal asymmetry, consider the movement of the wall: even if the light sphere stood still, the "sphere alone" system would be different at t_2 from what it was at t_1.

conserving mental interactions. Thus, the modal claim of P1' in OEC' is refuted and any *a priori* objection from energy conservation against dualism fails. I take it that this changes the discussion in two substantial ways:

1) Dualists can use what Brian Pitts (2018) dubs the 'conditionality answer': energy is conserved *on the condition that* no non-physical influence is acting. Thus, dualistic construals of mind-brain-interaction need no longer be restricted to energy-conserving accounts.
2) With energy conservation being shrunk to a factual (as opposed to modal) principle, the discussion can no longer rely on 'brain-free arguments' (Pitts 2018), i.e. arguments without consideration of the details of neuroscience, but is urged to consider empirical findings which alone can answer the question whether energy is conserved in brains or not.

Heeding point 2, I will now present an argument against physicalism which builds on the (empirically supported) assumption that energy is *not* conserved in brains.

3. Turning the Tables: the Objection from Energy Non-Conservation Against Physicalism

Physicalist philosophers generally assume that in brains, only physical causes are at work (which entails that energy is conserved). Thus Montero and Papineau:

> [I]t was the empirical evidence (…) that persuaded philosophers to be physicalists. Once mid-century physiological research had established that all physical effects had physical causes, even in bodies *and brains*, philosophers quickly figured out that general physicalism followed. (Montero and Papineau 2016, 188; italics added)

If Montero and Papineau know of experiments that have established *for brains* that "all physical effects have physical causes", they have not shared them with us. It is this unchecked assumption that I am going to contest in my argument 'objection from energy non-conservation' (OENC) against physicalism, which runs as follows:

<u>OENC:</u>

P1" Physicalism entails that energy is conserved in brains.

P2" Energy is probably not conserved in brains.

C" Therefore, physicalism is probably false.

The conclusion follows validly from the premises, so all that needs to be done is to explicate and defend the premises.

Premise 1

By physicalism I mean in this context the thesis that "every physical effect which has a cause has a physical cause". This is a 'weaker', less demanding version, in order to grant physicalism the maximum scope of possibilities to account for physical effects. The stronger variant has it that "every physical effect has a physical cause", thereby precluding some interpretations of quantum mechanics and the possibility of generic 'ontic chance', i.e. effects without cause. According to the weak version, physicalism claims that physical effects come about in either of two ways:

- They have a physical cause.
- They have no cause, i.e. they occur uncaused.

It is worth noting what "to occur uncaused" might mean. It might of course mean what I call 'ontic chance', that is, that the effect in question occurs 'ex nihilo'. It might, however, also mean that an effect is of quantum-mechanical nature (and I take it that this is what the definition is supposed to accommodate). For example, the quantum-mechanical decay of a single Radium atom is something that does not seem to have a cause. To be sure, its occurrence can (to some degree) be described by probability distributions, but it is hard to see a cause for it, unlike in the case where neutron bombardment triggers radioactive decay. With respect to P1, equating 'uncaused effect' with 'quantum effect' makes no difference, because quantum effects do conserve energy. However, if "to occur uncaused" is to mean 'ontic chance', things look different. Then there could be effects which occur 'ex nihilo', which entails that energy is not conserved. I do not think that any physicalist wishes to hold this. In fact, if a physicalist were prepared to accept ontic chance with respect to brain events, this would take them so close to dualism that one had to wonder if their continued rejection of it is but a mere prejudice.

In summary, it seems that a physicalist has two possibilities to account for physical effects, namely classical physical causes and quantum events, both of which conserve energy. I therefore take it that P1 is true.

Premise 2

Recall what energy non-conservation means in terms of physics. One could either describe it as "A system in which energy is not conserved has a Lagrangian in which the variable t figures" or as "A system in which energy is not conserved is asymmetrical with respect to time translation". As we saw in section 2, this might occur when a non-physical influence interacts with the system. In this case, the system's energy would increase (or decrease) without there being a physical cause for the energy rise (drop). Thus, what we are looking for is a scenario in which the energy in the brain increases or decreases without there being a physical cause for that change. I suggest that the ideal candidate for such a scenario is volitional action. By volitional actions I mean those body movements that the subject reports to have 'willed' and which occur independently of external stimuli[31]. Thus, the brain processes involved in those actions come as close as possible to being caused by a putative immaterial mind[32]. Let us now see what neuroscience tells us about the causal origins of volitional actions.

I take Haggard's (2008) [33] overview to reflect current knowledge of the neuro-causal picture of volitional action. It depicts the causal history of volitional actions roughly as follows:

Basal ganglia (BG) [34] ***→ prefrontal/frontopolar cortex (FPC)***[35] ***→ preSMA → SMA → primary motor cortex → spinal cord → muscles***

Importantly, actions triggered by external stimuli take a different path[36]. The neurobiological difference between externally triggered and 'pure' volitional actions is further corroborated by an activity increase in the SMA during the *mental exertion* of motor actions (Roland et al. 1980; Roland 1981)

[31] Cf. Haggard 2008.
[32] Reflexes mostly do not involve the brain and thus, *a fortiori*, they do not seem to involve the (conscious) mind either. Perceptually triggered actions seem to be good candidates for identifying an uninterrupted physical causal chain and thus conservation of energy.
[33] Haggard considers 'cue-free' studies, i.e. in which the participants were not told when to act. To be sure, some of those studies (e.g. Libet et al. 1983); Libet, Wright, and Gleason 1983) have been criticized for subtly 'nudging' people to act (see replies in Libet 1985), but, following Haggard, I take the results of the formally cue-free papers to be valid, not least because they widely concur regarding the big picture.
[34] See e.g. Picard and Strick 1996; Akkal, Dum, and Strick 2007.
[35] Soon et al. 2008
[36] Haggard 2008., 937; Brinkman and Porter 1979, 703-04

The study of brain processes is basically carried out by measuring *activity*[37]. Thus, the above causal chain can be considered as a sequence of activity increases in the respective brain regions. No matter which method is used to measure activity, the regional activity is taken to reflect activity of the neurons in that region[38], which in turn implies a change in energy. It is also important to emphasize that "activity increase of neurons" means an *increase in firing rate,* not a transition from a state of complete rest to a state of firing; in other words, neurons have a ground-state (or 'baseline'[39]) of (low-frequency) firing[40].

I will call the place where the chain of physiological causes leading to muscle contraction begins *in the brain* the 'neuro-causal origin' (NCO). The NCO might or might not itself be triggered by a cause outside the brain; that is one part of the empirical question before us. As it stands, the NCO of volitional actions is located in the BG; but even if it were located in a different brain region, there would always be an energy increase in that region. The central question we need to address is therefore "What makes neurons increase their firing rate?"

As indicated, a neuron can fire by itself without influence from other neurons (which constitutes the firing baseline). An increase in firing rate normally occurs when the neuron gets enough excitatory input from a presynaptic neuron. Consider the following figure:

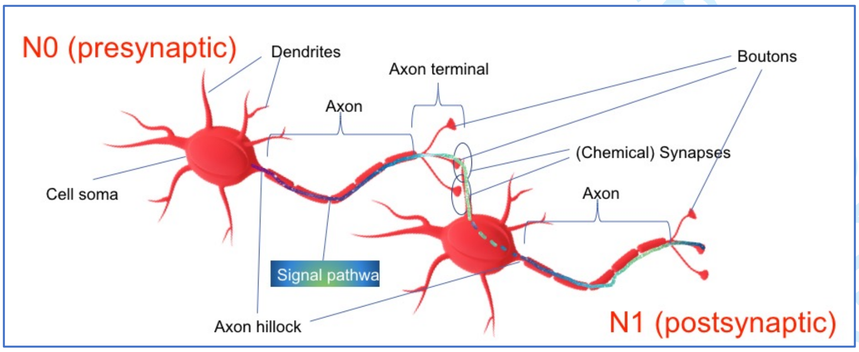

Figure 6: Two consecutive neurons (© Alin Christoph Cucu)

In this case, N1 gets excitatory input from N0. That means that N0's boutons release neurotransmitters into the synaptic cleft which then bind to N1's dendritic receptors:

[37] The activity measuring methods include electric potentials (e.g. Libet, Wright, and Gleason 1983; Libet et al. 1983; Deecke and Kornhuber 1978), regional cerebral blood flow (rCBF) (e.g. Roland et al. 1980; Roland 1981; Jahanshahi et al. 1995, 1995), firing rates of neurons ((Fried, Mukamel, and Kreiman 2011) (Brinkman and Porter 1979; J. Tanji and Kurata 1982)(Jun Tanji and Keisetsu 1994) and functional magnetic resonance imaging (fMRI) (e.g. Soon et al. 2008).
[38] For single neuron measurement, this is obvious.
[39] Fried, Mukamel, and Kreiman 2011
[40] See e.g. Stevens 1993

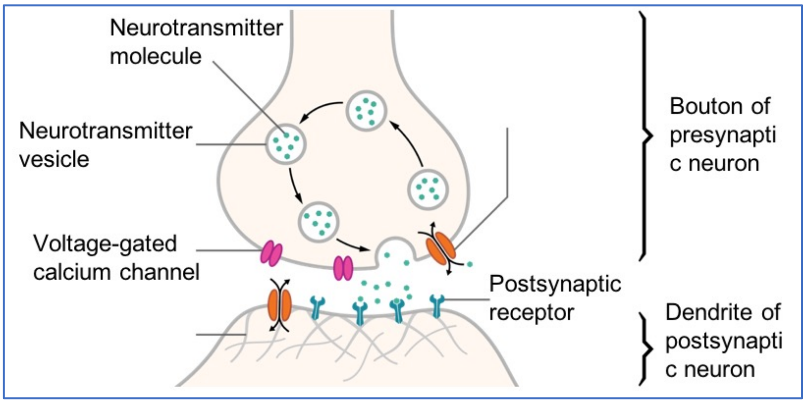

Figure 7: Synaptic communication (Thomas Splettstoesser – CC BY-SA 4.0; labels added)

The binding of the neurotransmitters triggers a so-called action potential (AP)[41] in N1, a current rapidly traveling down the axon caused by the opening of sodium and potassium (Na^+/K^+) channels. An AP in turn leads to the release of neurotransmitter molecules at the boutons (into the synapse toward N2 which is not on the picture). Thus, synaptic transmission works roughly as follows: N0 fires, releases neurotransmitters which bind to N1's dendritic receptors; N1 fires, releases NT and triggers N2 to fire; and so forth[42]. It is evident that this sequence of synaptic transmission cannot go back indefinitely; it must have a beginning, that is, a genuine 'neuron zero' or perhaps a number of such neurons[43] (equivalent to the NCO) which do the first firings (for simplicity's sake I shall hereafter talk only of the NCO in singular, as if it were a single neuron). What might make the NCO increase its firing rate?

A first natural suggestion is that the NCO is in fact triggered by a preceding neuron, which is linked to a causal chain that leads outside the body. One such scenario are stimulus-driven actions: in principle, there could be an uninterrupted physical causal chain from the external stimulus (e.g. a visual or tactile stimulus) through sensory cells and afferent nerves to the cortex, and from there via efferent nerves to the muscles. However, as indicated above, such actions take a different path through the brain; also, the studies underlying Haggard's overview are all 'cue-free' (see footnote 33). Hence, this possibility seems barred.

According to another hypothesis, the NCO is triggered by hormones coming from an endocrinal gland. I call this the *endocrinal hypothesis*. The endocrinal hypothesis perfectly retains energy conservation. However, I am not aware of any brain areas which could fit the endocrinal

---

[41] A current rapidly traveling down the axon.

[42] Of course, neurons do not just form such simple *chains*, but rather complex *networks*. But the sequentiality of synaptic transmission remains the same.

[43] There is the very real possibility that the causal origin consists of *one* neuron. Such 'command neurons' have been found in invertebrates (Stein 1978).

approach. The dopaminergic influence of the substantia nigra (SN) on the basal ganglia[44] unfortunately does not come into question as a candidate for triggering the NCO. The SN is not an endocrinal gland, but consists itself of neurons; also, the SN has afferences from the motor and premotor cortices, which means that while it (regulatorily[45]) influences cortical processes, it is itself influenced by the cortex. All this makes it a poor candidate for an NCO trigger. Of course, further research might find an endocrinal or similar candidate for an NCO trigger. However, it must be noted that endocrinal influences, being modulatory in nature, generally seem to be too slow for volitional actions[46].

A third option which clearly respects energy conservation are pacemaker cells. Those neurons regularly self-generate APs due to a cyclic mechanism of ion in- and outflux[47]. However, it is doubtful that such neurons can be found in the brain[48]; also, their activity is one of strict (though perhaps modifiable) regularity, which contradicts the idea of 'irregularly willed' voluntary actions. Again, if research were to find such neurons as the source of volitional actions, the physicalist could consider P2" refuted and physicalism vindicated.

The first two of the abovementioned proposals have in common that they rely on the regular receptor-mediated generation of APs. But in principle APs can be generated by other mechanisms, as seen in pacemaker cells. Those mechanisms all include proteins in some way or another. Consequently, the following hypotheses all involve conformational changes of some proteins, which basically requires energy expenditure[49]. One option is that sodium or potassium channels might open spontaneously or be caused to open in a deviating way by the binding of molecules, thereby triggering an AP[50]. By the same token, voltage-dependent Ca^{2+}(calcium) channels in the boutons might open without there having been a prior voltage change (i.e. AP) [51]. It is the calcium influx upon the opening of those channels that causes the release of the neurotransmitter vesicles from the bouton. Third, Ca^{2+} might be released from intracellular protein stores[52]. It would then have the same effect as extracellular calcium flowing in. Fourth,

44 cf. Haggard 2008, 936.
45 Pessiglione et al. 2006
46 Wilson 1999, 191-92; Hille 2001, ch. 20
47 Hille 2001, 147-49
48 In the human body they are known to exist in the heart.
49 For a good overview of research concerning protein conformational change in neurons see Wilson 1999.
50 To be sure, there are so-called ligand-gated sodium channels whose occurrence is, however, restricted to the neuromuscular junction (cf. Hammond 2015, ch. 6)
51 Normally those channels respond only to *large* membrane depolarizations (Hammond 2015, 151).
52 The intracellular stores are proteins located in the endoplasmatic reticulum, the calciosome, the mitochondria and the cytosol (the cytosolic stores are lightweight proteins like parvalbumin and calbindin) (ibd., 155). A release of calcium from there occurs normally upon an appropriate signal (e.g. the formation of inositol triphosphate) through

neurotransmitter vesicles might spontaneously be released from the axon terminal by exocytosis, which also requires the conformational change of some proteins[53]. How might those proteins be modified in ways that preserve energy? The following options come to mind:

1) Deviant ligand molecules (other than neurotransmitters) bind to the proteins
2) 'Outlier' molecules with kinetic energy far above average hit the proteins
3) Quantum effects are responsible for the spontaneous modification of the proteins

As to 1): There are indeed substances which activate sodium, potassium, and calcium channels[54], but all of them are pharmacologically active chemicals supplied from outside. As regards vesicular release, there do not seem to be any endogenous substances coming into question to trigger it[55]. The same holds true for calcium release from intracellular buffers.

Concerning option 2), statistical thermodynamics tells us that temperature is a measure for the *mean* kinetic energy of particles and that at any temperature, there are very few molecules far above/below that mean energy. Could not such 'outlier' particles, e.g. water molecules, be responsible for channel opening/vesicular release? They could in principle, but there are problems. First, the frequency of such events, given their low probability, seems to be insufficient to account for volitional action. Second, the approach seems much better suited for explaining the baseline firing rate, (which is clearly too low[56]); but if it explains the baseline, it cannot also explain the *increase* in firing rate.

Could the NCO be triggered quantum-mechanically (option 3)? Beck and Eccles (1992) calculated that the range of the Heisenberg uncertainty suffices for a spontaneous vesicle release to occur without energy expenditure. However, the validity of those calculations has been impugned[57]. But apart from that, the frequency of such quantum events seems too low; they, too, could better explain the baseline[58]. To fix the account, one might suggest a combination of

Ca-permeable channels. the proteins primarily serve as calcium-*binders* to reduce cytosolic Ca^{2+} (which is toxic in too high concentrations) (ibd., 51).

53 Südhof 1995

54 Examples include: Alkaloid-based toxins such as aconitine, batrachotoxin or brevetoxin for sodium channels (Hammond 2015, 68); diazoxide and minoxidil for potassium channels; and Bay K8644 and Ambroxol for calcium channels (Rang 2003, 60).

55 As an example of a vesicle-release-activating neurotoxin, latrotoxins present in black widow spiders cause *all* of the neuron's vesicles to release their neurotransmitters (Ushkaryov, Rohou, and Sugita 2008). This causes extreme pain and often death.

56 For example, in the absence of an AP, a spontaneous vesicle release in the frog neuromuscular junction is estimated to have a rate of 10^{-2}-10^{-3} times per second and release site, which means it occurs once every 100 to 1000 seconds. (Stevens 1993, 56)

57 Wilson (1999). In summary, Wilson's own calculations come to the result that within the confines of the Heisenberg uncertainty, the time for ion channels to stay open is far too short to account for sufficient ion influx (ibd., 196-199).

58 As Eccles (1994, ch. 4, 5) seems to suggest.

'outlier molecules' and quantum events: either outliers explain the baseline and quantum events the increase, or vice versa. The former version is implausible: how can a quantum cause have a greater and more frequent effect in the macroworld than a classical cause? The latter is *prima facie* more plausible, but it faces the intrinsic problems of the outlier hypothesis pointed out above. There is a last possibility for the physicalist not mentioned so far. It is the idea that energy is somehow *redistributed* within the brain at the time of the initiation of volitional actions. But we need not look too far to see a fatal problem with this proposal: such redistribution violates (linear) momentum conservation, because clearly some particles would have to change their direction 'spontaneously', without there being a physical cause for that change. The physicalist would then trade energy conservation for momentum conservation, which still leaves him with a big explanatory gap (as to what caused the violation of momentum conservation).

In summary, it seems that there is no satisfactory energy-conserving account of the NCO. Therefore, I take it that as it stands, it is much more probable than not that energy is *not* conserved at the initiation of volitional actions. Therefore, premise 2 seems true and OENC succeeds.

Conclusion

According to modern physics, the modal, a priori version of OEC (see section 1) does not even get off the ground, because energy and momentum conservation hold only *conditionally* (see section 2).

Empirically, volitional actions begin with a local energy increase for which it seems very difficult to find a physical cause such that energy is conserved. But if energy is not conserved in brains, then physicalism, which entails energy conservation, is false. Physicalists have long thought their insistence on energy conservation to be a "perfect pass" against dualism. If, however, my arguments are correct, this perfect pass turns out to be a colossal own goal.

References

Akkal, D., R. P. Dum, and P. L. Strick. 2007. "Supplementary Motor Area and Presupplementary Motor Area: Targets of Basal Ganglia and Cerebellar Output." *Journal of Neuroscience* 27: 10659–73.

Beck, Friedrich, and John C. Eccles. 1992. "Quantum Aspects of Brain Activity and the Role of Consciousness." *Proceedings of the National Academy of Science USA* 89: 11357–61.

Brinkman, C., and R. Porter. 1979. "Supplementary Motor Area in the Monkey: Activity of Neurons during Performance of a Learned Motor Task." *Journal of Neurophysiology* 42 (3): 681–709. https://doi.org/10.1152/jn.1979.42.3.681.

Broad, C.D. 1937. *The Mind and Its Place in Nature*. London: Kegan Paul, Trench, Trubner and Co.

Collins, Robin. 2008. "Modern Physics And The Energy-Conservation Objection To Mind-Body Dualism." *American Philosophical Quarterly* 45 (1): 31–42.
———. 2011. "A Scientific Case for the Soul." In *The Soul Hypothesis*, edited by Marc Baker and Stewart Goetz, 222–46. continuum.
Deecke, L., and H. H. Kornhuber. 1978. "An Electrical Sign of Participation of the Mesial 'supplementary' Motor Cortex in Human Voluntary Finger Movement." *Brain Research* 159 (2): 473–76. https://doi.org/10.1016/0006-8993(78)90561-9.
Dennett, Daniel C. 1991. *Consciousness Explained*. Penguin Books.
Dilley, F.B. 2004. "Taking Consciousness Seriously: A Defense of Cartesian Dualism." *International Journal for Philosophy of Religion* 55: 135–53.
Ducasse, C. 1960. "In Defense of Dualism." In *Dimensions of Mind: A Symposium*, edited by S. Hook, 85–90. New York: New York University Press.
Eccles, John C. 1994. *How the Self Controls Its Brain*. Springer.
Fried, Itzhak, Roy Mukamel, and Gabriel Kreiman. 2011. "Internally Generated Preactivation of Single Neurons in Human Medial Frontal Cortex Predicts Volition." *Neuron* 69: 548–62.
Garber, Daniel. 1983. "Mind, Body and the Laws of Nature in Descartes and Leibniz." *Midwest Studies In Philosophy* 8 (1): 105–33. https://doi.org/10.1111/j.1475-4975.1983.tb00462.x.
Gibb, Sophie. 2010. "Closure Principles and the Laws of Conservation of Energy and Momentum." *Dialectica* 64: 363–84.
Goldstein, Herbert. 1980. *Classical Mechanics, 2nd Ed.:2nd (Second) Edition*. 8948th edition. Addison-Wesley.
Haggard, Patrick. 2008. "Human Volition: Towards a Neuroscience of Will." *Nature Reviews Neuroscience* 9 (12): 934–46. https://doi.org/10.1038/nrn2497.
Hammond, Constance. 2015. *Cellular and Molecular Neurophysiology*. 4 edition. Amsterdam ; Boston: Academic Press.
Hart, W.D. 1994. "Dualism." In *A Companion to the Philosophy of Mind*, edited by S. Guttenplan, Blackwell Companions to Philosophy, 265–69. Oxford: Blackwell.
Hille, Bertil. 2001. *Ion Channels of Excitable Membranes*. 3rd ed. Sunderland, MA: Sinauer Associates.
Jahanshahi, Marjan, I. Harri Jenkins, Richard G. Brown, C. David Marsden, Richard E. Passingham, and David J. Brooks. 1995. "Self-Initiated versus Externally Triggered MovementsI. An Investigation Using Measurement of Regional Cerebral Blood Flow with PET and Movement-Related Potentials in Normal and Parkinson's Disease Subjects." *Brain* 118 (4): 913–33. https://doi.org/10.1093/brain/118.4.913.
Larmer, Robert. 2009. "Divine Agency And The Principle Of The Conservation Of Energy." *Zygon* 44 (3): 543–57.
———. 2014. "Divine Intervention and the Conservation of Energy: A Reply to Evan Fales." *International Journal for Philosophy of Religion* 75 (1): 27–38.
Leibniz, G.W. 1985. *Theodicy: Essays on the Goodness of God and the Freedom of Man and the Origin of Evil.* Translated by E.M. Huggard. La Salle, Illinois: Open Court.
Libet, B., C. A. Gleason, E. W. Wright, and D. K. Pearl. 1983. "Time of Conscious Intention to Act in Relation to Onset of Cerebral Activity (Readiness-Potential). The Unconscious Initiation of a Freely Voluntary Act." *Brain : A Journal of Neurology* 106 (Pt 3) (September): 623–42. https://doi.org/10.1093/brain/106.3.623.
Libet, Benjamin. 1985. "Unconscious Cerebral Initiative and the Role of Conscious Will in Voluntary Action." *Behavioral and Brain Sciences* 8 (4): 529–39. https://doi.org/10.1017/S0140525X00044903.
Libet, Benjamin, E. Wright, and C. Gleason. 1983. "Preparation – Or Intention-to-Act, in Relation to Pre-Event Potentials Recorded at the Vertex." *Electroenceph. And Clin. Nerophysiology* 56: 367–372.
Lowe, Jonathan. 1992. "The Problem Of Psychophysical Causation." *Australasian Journal of Philosophy* 70 (3): 263–76.

Mcginn, Colin. 2000. *The Mysterious Flame: Conscious Minds in a Material World*. New Ed. New York, NY: Basic Books.

Meixner, Uwe. 2008. "New Perspectives for a Dualistic Conception of Mental Causation." *Journal of Consciousness Studies* 15 (1): 17–38.

Montero, Barbara Gail, and David Papineau. 2016. "Naturalism and Physicalism." In *The Blackwell Companion to Naturalism*, edited by Kelly James Clark, 1st ed., Ch. 13. John Wiley & Sons.

Noether, E. 1918. "Invariante Variationsprobleme." *Nachrichten von Der Gesellschaft Der Wissenschaften Zu Göttingen, Mathematisch-Physikalische Klasse*, 235–57.

Papineau, David. 2000. "The Rise of Physicalism." In *The Proper Ambition of Science*, edited by M. W. F. Stone and J. Wolff, 174–208. New York: Routledge.

Pessiglione, M. et al. 2006. "Dopamine-Dependent Prediction Errors Underpin Reward-Seeking Behaviour in Humans." *Nature* 442: 1042–45.

Picard, N., and P. L. Strick. 1996. "Motor Areas of the Medial Wall: A Review of Their Location and Functional Activation." *Cereb. Cortex*, no. 6: 342–53.

Pitts, Brian. 2018. "Conservation Laws and the Philosophy of Mind: Opening the Black Box, Finding a Mirror." *Forthcoming*.

Pitts, J. Brian. 2004a. "Has Robert Gentry Refuted Big Bang Cosmology? On Energy Conservation and Cosmic Expansion." *Perspectives on Science and Christian Faith* 56 (4): 260–65.

———. 2004b. "Reply to Gentry on Cosmological Energy Conservation and Cosmic Expansion." *Perspectives on Science and Christian Faith* 56 (4): 278–84.

———. 2009. "Nonexistence of Humphreys' 'Volume Cooling' for Terrestrial Heat Disposal by Cosmic Expansion." *Perspectives on Science and Christian Faith* 61 (1): 23–28.

Plantinga, Alvin. 2007. "Materialism and Christian Belief." In *Persons: Human and Divine*, edited by Peter Van Inwagen and Dean Zimmerman, 99–141. Oxford University Press.

Popper, Karl R., and John C. Eccles. 1977. *The Self and Its Brain: An Argument for Interactionism*. Springer.

Rang, H. P. 2003. *Pharmacology*. Edinburgh: Churchill Livingstone.

Roland, P. E. 1981. "Somatotopical Tuning of Postcentral Gyrus during Focal Attention in Man. A Regional Cerebral Blood Flow Study." *Journal of Neurophysiology* 46 (4): 744–54. https://doi.org/10.1152/jn.1981.46.4.744.

Roland, P. E., B. Larsen, N. A. Lassen, and E. Skinhøj. 1980. "Supplementary Motor Area and Other Cortical Areas in Organization of Voluntary Movements in Man." *Journal of Neurophysiology* 43 (1): 118–36. https://doi.org/10.1152/jn.1980.43.1.118.

Searle, John R. 2004. *Mind: A Brief Introduction*. Oxford University Press.

Soon, Chun Siong, Marcel Brass, Hans-Jochen Heinze, and John-Dylan Haynes. 2008. "Unconscious Determinants of Free Decisions in the Human Brain." *Nature Neuroscience* 11 (5): 543–45. https://doi.org/10.1038/nn.2112.

Stein, P S G. 1978. "Motor Systems, with Specific Reference to the Control of Locomotion." *Annual Review of Neuroscience* 1 (1): 61–81. https://doi.org/10.1146/annurev.ne.01.030178.000425.

Stevens, C. F. 1993. "Quantal Release of Neurotransmitter and Long-Term Potentiation." *Cell* 72 Suppl (January): 55–63. https://doi.org/10.1016/S0092-8674(05)80028-5.

Südhof, Thomas C. 1995. "The Synaptic Vesicle Cycle: A Cascade of Protein–Protein Interactions." *Nature* 375 (6533): 645–53. https://doi.org/10.1038/375645a0.

Tanji, J., and K. Kurata. 1982. "Comparison of Movement-Related Activity in Two Cortical Motor Areas of Primates." *Journal of Neurophysiology* 48 (3): 633–53. https://doi.org/10.1152/jn.1982.48.3.633.

Tanji, Jun, and Shima Keisetsu. 1994. "Role for Supplementary Motor Area Cells in Planning Several Moevements Ahead." *Nature* 371: 413–16.

Ushkaryov, Yuri A., Alexis Rohou, and Shuzo Sugita. 2008. "α-Latrotoxin and Its Receptors." In *Pharmacology of Neurotransmitter Release*, edited by Thomas C. Südhof and Klaus Starke, 171–206. Handbook of Experimental Pharmacology. Berlin, Heidelberg: Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-540-74805-2_7.

Von Wachter, Daniel. forthcoming. "The Principle of the Causal Openness of the Physical." *Organon F*.

Westphal, Jonathan. 2016. *The Mind–Body Problem*. Cambridge, Massachusetts: MIT Press.

White, Ben. 2016. "Conservation Laws And Interactionist Dualism." *The Philosophical Quarterly* 00 (0).

Wilson, David. 1999. "Mind–Brain Interaction and Violation of Physical Laws." *Journal of Consciousness Studies* 6 (8–9): 185–200.

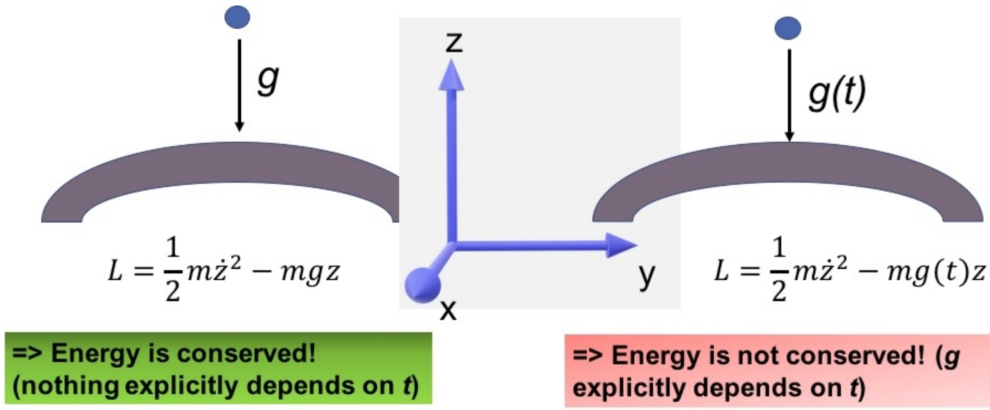

Figure 1: Particle in one-dimensional gravitational field (© Alin Christoph Cucu)

319x131mm (72 x 72 DPI)

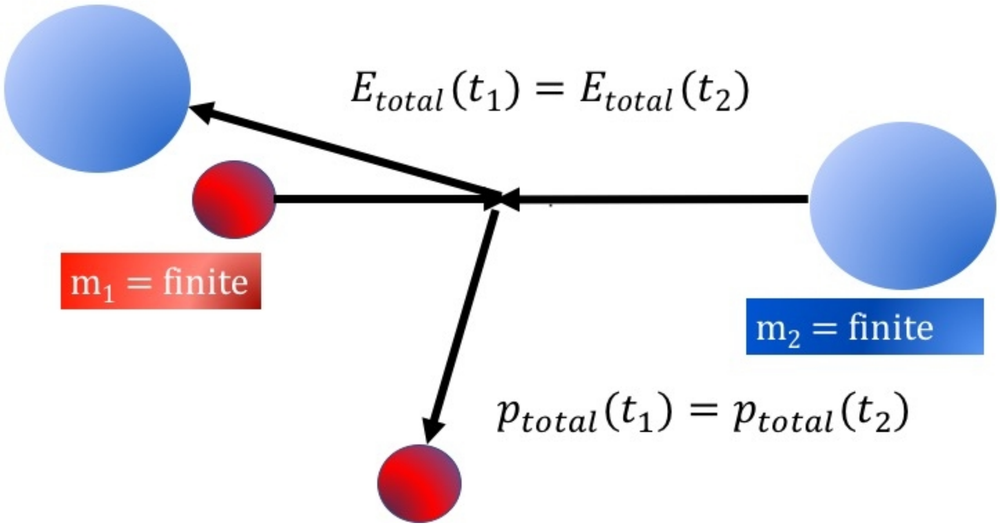

Figure 2: Collision of two spheres with finite mass (© Alin Christoph Cucu)

255x133mm (72 x 72 DPI)

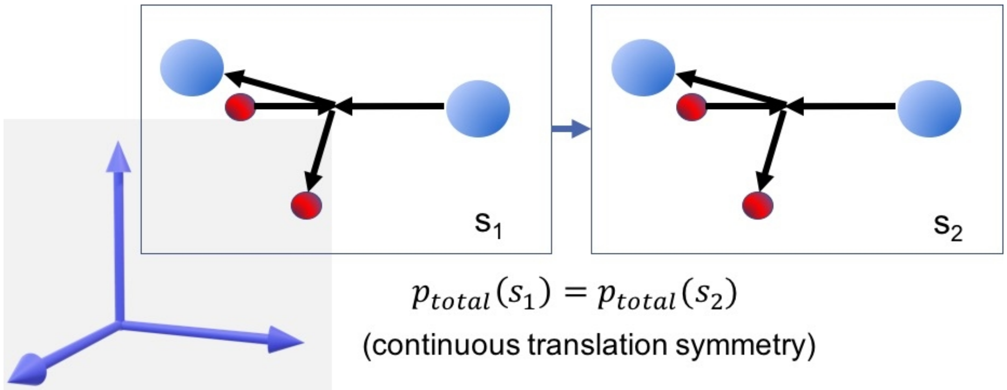

Figure 3: Space translation symmetry of the two-spheres system

326x126mm (72 x 72 DPI)

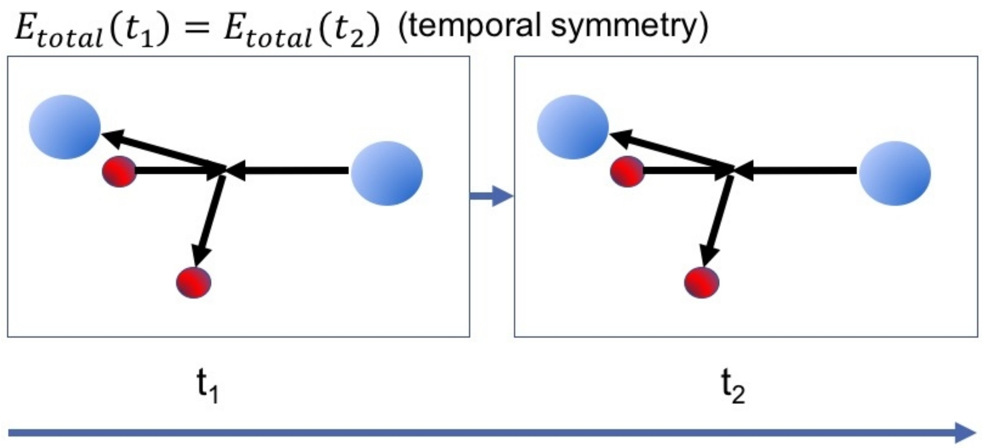

Figure 4: Time translation symmetry of the two-spheres system (© Alin Christoph Cucu)

289x134mm (72 x 72 DPI)

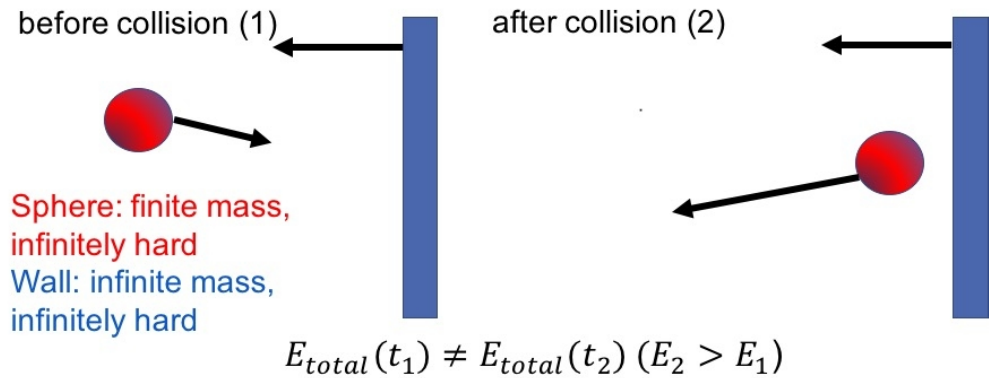

Figure 5: 'Toy model' of mental interaction with a massive object (© Alin Christoph Cucu)

311x126mm (72 x 72 DPI)

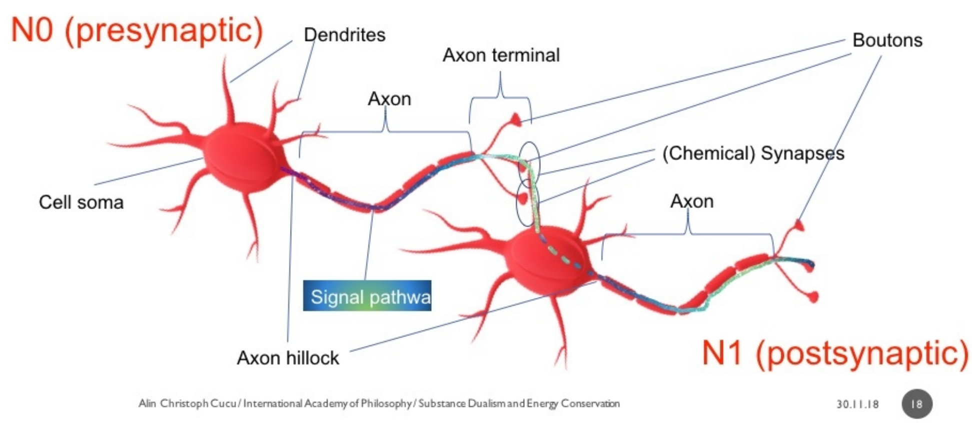

Figure 6: Two consecutive neurons (© Alin Christoph Cucu)

330x156mm (72 x 72 DPI)

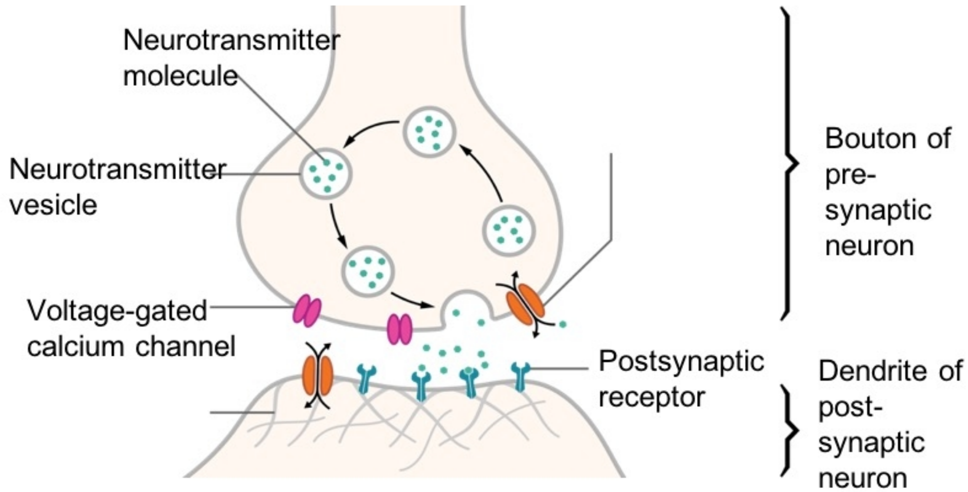

Figure 7: Synaptic communication (Thomas Splettstoesser – CC BY-SA 4.0; labels added)

286x145mm (72 x 72 DPI)