Turning the Tables: How Neuroscience Supports Interactive Dualism

Alin Cucu
Department of Philosophy
University of Lausanne, Switzerland

Abstract

Physicalists typically believe that neurophysiology has refuted the thesis that non-physical minds can interact with the brain. In this paper, I argue that it is precisely a closer look at the neurophysiology of volitional actions that suggests otherwise. I start with a clarification of how the present inquiry relates to the main argument for physicalism, and how the most common alternative views relate to the findings of my study. I then give a brief overview of the neurophysiological research about volitional actions, finding that there is no research specifically directed at the pertinent question. I proceed by pointing out what it would take for a complete physical explanation of volitional actions to be true: namely a complete physical explanation of the increase in the firing rate of the neurons with which the sequence leading up to volitional actions starts. Since no dedicated research about this question is available, I offer a study of the known mechanisms of neuronal excitation as a substitute, finding that there is no plausible biochemical or physical mechanism that could explain the causal initiation of volitional actions – at least none that upholds energy conservation. But non-conservation is precisely what interactive dualism, in its most plausible version, predicts. Thus, rather than buttressing physicalism, our empirical knowledge of volitional actions points toward interactive dualism.

1. Introduction

My main target in this paper is the philosophical position of physicalism. Typically, physicalism is the thesis that everything is physical, even things that at first glance seem non-physical (like minds). It arrives at this rather strong ontological claim by way of the premise that the physical world is causally closed, which in turn rests heavily on the assumption that the sciences show empirically that all physical effects, even in human bodies, have physical causes. My aim here is to challenge this latter assumption. If my arguments are successful, the scientific foundation of physicalism receives a significant blow. In this introductory section I wish to first clarify how the physicalist reasoning goes, and then offer

a brief mapping of the alternatives to physicalism, all in relation to the subsequent neuroscientific discussion.

The leading argument for physicalism uses as its crucial (and first) premise the causal closure of the physical[1] (see Stoljar 2021, Sec. 6), henceforth referred to as CCP. CCP, in a common version (Kim 1998, p. 40), has it that if

> ... you pick any physical event and trace out its causal ancestry or posterity, that will never take you outside the physical domain. That is, no causal chain will ever cross the boundary between the physical and the nonphysical.

Note that Kim's version of causal closure already contains a rejection of causal overdetermination: bodily actions can never be caused by both physical and mental events, since "no causal chain will ever cross the boundary between the physical and the nonphysical".[2] That by itself is still compatible with the existence of *sui generis* mental events (or substances), albeit ones that do not cause any physical events. However, the resulting position – epiphenomenalism – is unappealing, because it so clearly contradicts our common-sense belief that it is us who mentally cause physical movements in our bodies; Jerry Fodor (1983) even famously wrote it would be "the end of the world".[3] This is why physicalists typically seek to save the causal efficacy of mental events by making them physical (see Kim 1998).[4]

[1] By "physical" I here mean the standard definition: the physical is everything that is (i) described by physics and (ii) not fundamentally mental. This includes parts and processes of living bodies. Neuronal processes are physical in this sense. The dichotomy here is between the physical and the mental, not the physical and the chemical or biological.

[2] Other versions of CCP require the rejection of causal overdetermination (also called the exclusion principle) as an auxiliary premise in the argument for physicalism. Thus David Papineau's version, which he calls the "completeness" of physics: "All physical effects are fully determined by law by a purely physical prior history" (Papineau 2001, Sec. 4). The "full determination" Papineau speaks of is arguably equivalent to sufficient causation, which still leaves room for causal overdetermination. Moreover, the "sufficient causation" construal may run into problems concerning quantum mechanics, which is at least consistent with indeterminism at the micro-physical level. One typical physicalist rejoinder that physical causes are sufficient for the probabilities of the effects may or may not be successful. On the other hand, retreating to the position that quantum indeterminism cancels out at the meso- or macroscopic level is an empirical issue, which I argue in this paper is far from settled. I am indebted to an anonymous reviewer for pointing these things out to me.

rejoinder - that…effects - …

[3] One could obtain the rejection of epiphenomenalism also via a different route, namely by the application of the ontological parsimony enshrined in the Eleatic principle, which holds that entities should only be added to the ontology if they make a causal difference.

[4] Unfortunately, physicalists often describe mental events as "supervening" on physical events. The problem is that the supervenience of the mental on the physical is compatible with mental events being of an ontological *sui generis* order; in fact, in

One may ask, though, why accept CCP? After all, there does not seem to be any compelling *a priori* reason to embrace it – which even physicalists admit. As it turns out, it is chiefly *empirical* evidence that physicalists invoke as justification (Montero and Papineau 2016, p. 188):

> It was the empirical evidence for causal closure 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's and Papineau's point is that physicalism is not an "armchair" philosophy, but empirically grounded. More specifically, they claim that physiological research has shown that all brain events, being physical, have physical causes. How this is supposed to imply the general metaphysical thesis of CCP, and whether there can possibly be a convincing argument leading from empirical science to CCP,[5] need not bother us here. Let us assume for the sake of argument that the physicalist reasoning is valid. If it could be shown that the empirical premise in the physicalist argument is false, the argument would be considerably weakened, since CCP would lose its strongest supporting premise.

In this article, I therefore wish to challenge the aforementioned physicalist assumption that physiological research has established, even in brains, that all physical effects have physical causes. In fact, I even wish to turn the tables on it: it is by close inspection of the neurophysiology of the relevant phenomenon – volitional actions – that one can come to see that the involved neuronal processes, for all we know, *lack* a complete physical explanation, at least one that upholds energy conservation. But non-conservation, I will argue, clearly points to the involvement of a non-physical cause. I shall attempt this as follows: first, I will give an overview of what is known about the neurophysiology of volitional actions. Then, I will show that, given what we know, a complete (i.e., energy-conserving) physical explanation of volitional actions seems *in principle* unavailable.

Before we start the investigation in earnest, I need to say something about the alternatives to physicalism in terms of the empirical results obtained. If my conclusion is true, then what is called for is an ontology which both conceives of the mind as non-physical and predicts interactions of the non-physical mind with the brain. According to the conception of mind-body interaction under discussion here, those interactions are counterfactually robust,[6] in order to make a causal difference: in short,

my view talk of supervenience makes much more sense on property dualism than on physicalism. Physicalists should, in accordance with their general ontological thesis, say that mental events are physical events. "are" in italics

[5] In fact, I believe there cannot; see for example von Wachter (2019).

[6] I am aware that there are versions of interactive dualism on which mental influence is not counterfactually robust, see e.g. Lowe (2006).

had the mind not interacted, things in the brain would have been different. In conformity with the result of my investigation (see Secs. 3.4 and 4), the interactions would also have to *not* conserve energy.[7]

A theory that definitely predicts such a causal contribution of the mind is interactive dualism, where the non-physical mind is construed in terms of either mental properties,[8] or a mental substance. Although my own position is that of substance dualism (though with an important qualification[9]), nothing turns on the distinction between substance and property dualism for the sake of this paper.

Interactive (substance) dualism has a venerable history, already appearing in Plato, where the soul[10] is a self-moving entity that has the power to move the body, and which is indispensable for the explanation of actions (see Goetz and Taliaferro 2011, pp. 13ff). Its best-known (early) modern proponent is René Descartes. In recent years, interactive dualism has been defended by thinkers such as Charles Taliaferro (1994), Richard Swinburne (1997, 2013, 2019), Moreland and Rae (2000), Robin Collins (2008, 2011a, 2011b), William Hasker (1999), E.J. Lowe (1992, 2006), Uwe Meixner (2004, 2008, 2019), Rickabaugh and Buras (2017), Ralph Stefan Weir (2023), Joshua Farris (2016, 2023), Alin Christoph Cucu and Brian Pitts (2019) Pitts (2019), Cucu (2022), as well as John Eccles and Friedrich Beck, see (Eccles 1994, Beck and Eccles 1992, Beck 1996, 2008). (Cucu and Pitts 2019, Pitts 2019, Cucu (2022))

One big divide between those views runs between accounts that seek to make interaction energy-conserving (Swinburne, Meixner, Collins in his 2011b, Eccles and Beck)[11] and accounts that accept energy non-conservation (e.g. Taliaferro, Collins 2008, Cucu and Pitts). As will become clear in due course, this paper goes along with the latter approach.

The two other main non-physicalist contenders for a theory of mind are hylomorphism and panpsychism.[12] It is beyond the scope of this paper

[7] Counterfactual robustness and energy non-conservation plausibly go together, but perhaps not necessarily. If some quantum-mechanics based account of mental interaction is true, then mental interaction could both make a causal difference and conserve energy. There is, however, reason to be skeptical about the exploitation of quantum indeterminacies being energy-conserving (see Cucu 2020).

[8] I am not aware of any contemporary defender of interactive property dualism. Chalmers' and McQueens' "interactive" property dualism (Chalmers and McQueen 2022) does not qualify as an interactive account in the sense specified, because its adherence to supervenience and causal closure renders interaction counterfactually unstable.

[9] My own version is Thomistic-like dualism (see Moreland 2018), according to which the body is not a substance in its own right, but a mode or inseparable part of the soul-organism.

[10] Plato uses the term "soul" for that which most contemporary dualists call "mind".

[11] Although momentum conservation is physically as momentous (pun intended) as energy conservation, it is barely ever addressed by philosophers dealing with interactionism.

[12] For a hylomorphic account concerned with mental causation see Owen (2020). For an exposition of current debates around panpsychism see Brüntrup and Jaskolla (2017).

to ascertain whether on those accounts mind-body interaction is counterfactually robust; to the extent to which it is, they come into question as alternatives to interactive dualism. At any rate, of all the non-physicalist alternatives interactive dualism is most clearly in contrast to physicalism. For the sake of simplicity, I will therefore speak of interactive dualism (which on another occasion I would explicitly defend against rival views) as representative of all non-physicalist, counterfactually robust accounts. This should not be taken to imply that the other theories do not constitute viable alternatives to physicalism.

2. The Neurophysiology of Volitional Actions

First, it is important to make clear what I understand by "volitional (or voluntary) action". I use the term here in its neurophysiological sense, where it refers to those body movements that the subject reports to have "willed" and which occur independently of external stimuli. This is also the way the reviews of Haggard (2008) and Fried *et al.* (2017) use the term. To be sure, there are other working definitions of voluntary actions circulating, but, as Fried *et al.* (2017, p. 10843) point out, "the canonical form of voluntary action is self-initiated action (SIA), which we define as voluntary action initiated without any sensory cue". This latter addition to our everyday concept of volitional action (which has it that a volitional action is an action brought about "consciously" or "intentionally" as opposed to reflexively) is necessary because it exclusively considers the causal contribution of the agent and rules out external factors like sense stimuli.

It should be noted that in terms of the framing as a neural event (as opposed to framing as an event involving the person and her volitions), there is an ongoing debate among neurophysiologists as to how to understand voluntary actions. Consider the following quotation from the recent review by Fried *et al.* (2017, p. 10843, italics added, and RP stands for readiness potential):

> If voluntary action is reliably linked to a specific process in the brain, then in principle "voluntary action" could be defined as any action that is preceded by and caused by such a neural state. But if the RP does not reflect such a state, then either we must look for another putative correlate or else consider that "voluntary action" may not be a well-defined category in terms of neural phenomena. Rather, it may be a perceptual attribution: an action is voluntary if and only if it is perceived by the agent as being voluntary. This perception presumably must have some neural basis (*unless one is a dualist*).

The authors suggest that voluntary action might be a perceptual category – that it happens neuronally and is concomitantly *perceived* rather

than *generated* by the agent. But the last qualification in parentheses makes it clear that there is a way open to dualists to interpret things dualistically.

Now, which brain areas are involved in the generation of volitional actions? Many readers will be familiar with the work of Benjamin Libet (Libet *et al.* 1983a,b), which was itself based on previous work by Deecke and Kornhuber (1978). Those latter investigations found the SMA (supplementary motor area) to be reliably involved in voluntary actions, since the occurrence of a readiness potential (RP) in the SMA correlated consistently[13] with the occurrence of the mental event of volition as well as the onset of muscle contraction. Since then, further research has been conducted that has allowed the sequence of events leading to muscle contraction to be traced back even further.

The following picture serves as demonstration that there is a sequence of events *sans* external sense stimuli (to the degree to which sense stimuli can be ruled out in laboratory conditions) leading up to volitional action. In fact, it has been empirically confirmed that actions triggered by external stimuli take a different path (Haggard 2008, p. 937, Brinkman and Porter 1979, pp. 703f). This means that when we speak of volitional actions we refer, neurobiologically, to a specific neuronal pathway free of external stimuli and cannot invoke the other, stimulus-induced, pathway for explaining volitional actions. The neurobiological difference between externally triggered and "pure" volitional actions is further buttressed by experiments showing that an activity increase in the SMA takes place during the *mental exercise* of motor actions (Roland *et al.* 1980, Roland 1981).

I take Haggard's (2008) overview to reflect current knowledge about the neurophysiology involved in volitional actions.[14] In a nutshell, and with the caveat that the following sequence basically represents a temporal succession which may or may not be causal, the order of activated brain regions is as follows:

> **basal ganglia (BG)**[15] **→ prefrontal/frontopolar cortex (FPC)**[16] **→ preSMA → SMA → primary motor cortex → spinal cord → muscles**

All areas in this chain (except muscles, of course) consist of neurons. Thus, when we speak of activity increases in brain regions, we really speak of

[13] The moot point about the Libet experiments was of course not whether there was correlation, but the temporal order of RP and volition.

[14] Ironically, one conclusion that Haggard draws from his investigations is that the dualistic picture of volitional action should be abandoned for a purely physicalist one with reflexes at one end of the spectrum and volitional actions at the other end.

[15] See e.g. Picard and Strick 1996, Akkal *et al.* 2007.

[16] See Soon *et al.* 2008

activity increases of neurons.[17] It is also important to emphasize that "activity increase" 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" (Fried *et al.* 2011) of (low-frequency) firing (see e.g. Stevens 1993) that increases when those neurons become activated.

closing bracket missing

Now, according to Montero and Papineau, a *complete* physical explanation for volitional actions should be available. However, the above picture isn't one: one can still meaningfully ask what makes the neurons of the basal ganglia fire (if indeed the chain of events starts there). At least to my knowledge, no research exists that is specifically dedicated to answering the question of a complete physical explanation of volitional actions. This alone considerably weakens the physicalist claim that physiology already *has* found physical explanations of *all* bodily functions. A typical physicalist rejoinder at this point is that physiological research has found no *anomalies* in brains (at least that is what I heard many times at conferences), but mental interactions would be such anomalies, so we are not justified in assuming such interactions to occur. But dualists can and should reply that interactions are arguably very "gentle", i.e. they don't make a *huge* difference to how things would have been without them. After all, minds interact with complex biological systems already teeming with activity. Interactions thus do not have to move biological structures from a state of complete rest but rather slightly alter the trajectories of systems already in motion.[18]

That still leaves us with nothing more but the *possibility* for interactive dualism to be true. It might be that a complete physical explanation of volitional actions will be found one day, which would render physicalism true; or it could be that at least some brain events require a non-physical explanation, in which case interactive dualism (or another non-physicalist view) would be true. Without dedicated empirical research there is no way of adjudicating between those two possibilities. Fortunately, we can advance the discussion even without such empirical investigations. The key is to reformulate the question and ask if there can *in principle* be a

[17] The activity measuring methods include electric potentials (e.g. Libet *et al.* 1983a,b, Deecke and Kornhuber 1978), regional cerebral blood flow (e.g. Roland *et al.* 1980, Roland 1981, Jahanshahi *et al.* 1995), firing rates of neurons (Fried *et al.* 2011, Tanji and Keisetsu 1994, Brinkman and Porter 1979, Tanji and Kurata 1982) and functional magnetic resonance imaging (e.g. Soon *et al.* 2008).

[18] In an unpublished manuscript Ralph Weir suggests that we should assume that if there are nonphysical forces acting in the brain, they will adhere to a "law of least action" according to which, when a change occurs in nature, the quantity of action necessary for the change is the least possible. After all, physical forces adhere to such a law, and so theoretical simplicity suggests that nonphysical forces will do so too. If so, then it is not just that the mind need not make a big difference to the dynamics of the brain, but we should expect that the differences it makes will be minimal.

complete physical explanation of volitional actions, given our best knowledge of basic neuronal functioning. This line of thought will take us to the workings of the smallest functional unit of the brain, the neuron. The question then becomes: which physical explanations are available *in principle* for increases in neuronal firing rate? To this we turn next.

3. In Search of a Complete Microphysical Explanation for the "Neuro-Causal Origin"

Suppose, for the sake of argument, that there is a complete physical explanation for volitional actions. As already mentioned, we do not know its details – research is missing here. What I therefore do in this Section is seek to construct a possible complete physical explanation out of what we *do* know about the functioning of the brain and brain cells in particular.

Brain activity consists of the activity of neurons, which in turn is constituted by their "firing" (more precisely, their issuing action potentials). That is common lore. Now, if there is a complete physical explanation of volitional actions, then the neuronal activity that ultimately leads to muscle contraction must be entirely explainable in physical terms. That explanation can either make reference exclusively to brain cells or include structures situated outside the brain. In any event, there will be neurons that fire first and thus set in motion the sequence of events that ends up in some motor movement. I shall call the first brain cell – or cells – lying in the sequence of events leading up to volitional muscle contraction the *neuro-causal origin* (NCO). Now, how can it come about that an NCO neuron fires (or, more precisely, increases its firing rate)? [19]

Replace "That" by "Such an"

The standard way for any neuron to increase its firing rate is when it receives sufficient excitatory input from a presynaptic neuron. As shown in Fig. 1, neuron 1 (N1) receives excitatory input from neuron 0 (N0). This means that N0's boutons release neurotransmitters into the synaptic cleft which then bind to N1's dendritic (postsynaptic) receptors. The binding of the neurotransmitters triggers a so-called action potential (AP) in N1, an electric current rapidly traveling down the axon caused by the opening of sodium and potassium (Na^+/K^+) channels. An AP in turn leads, mediated by the opening of voltage-gated calcium channels in the boutons (Fig. 2) to the release of neurotransmitter molecules at the boutons (into the synapse facing another neuron N2, which is not on the picture). Thus, synaptic transmission works roughly as follows: N0 fires and releases neurotransmitters which bind to N1's dendritic receptors; N1 fires, releases neurotransmitters and makes N2 fire; and so forth.[20]

[19] For ease of expression, I speak of "neuron" in the singular. The NCO might of course consist of many neurons, plausibly connected in a network-/circuit-like manner.

[20] Of course, neurons do not just form such simple chains, but rather complex net-

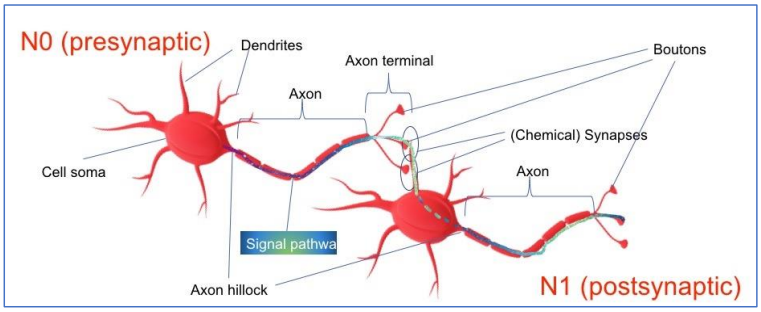

Figure 1: Schematic illustration of two consecutive neurons (© Alin Christoph Cucu)

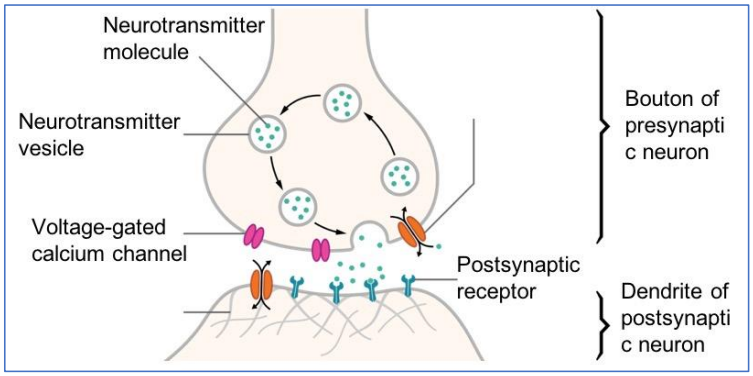

Figure 2: Details of synaptic communication (Thomas Splettstoesser – *CC BY-SA 4.0*; labels added)

It is evident that this sequence of synaptic transmission cannot reach back indefinitely; it stands to reason that it must have a beginning, that is, a genuine "neuron zero" or perhaps several such neurons (equivalent to the NCO) which do the first firings.[21] One might however object here that, since the brain is in constant activity, we should conceive of the sequence of neuronal events more as "circular" than linear, in the sense that neurons in the brain constantly stimulate each other; then there would be, in accordance with the image of a circle, no identifiable beginning.

The proposal by Schurger *et al.* (2012) that the continual background ebb and flow of brain activity influences the onset of voluntary actions seems to go in that direction. However, as plausible as that picture is, it

works where one neuron is connected to up to 10,000 other neurons. But the sequentiality of synaptic transmission remains the same.

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

is beside the point. The reason is that even with a "circular" sequence of events there are clear, measurable differences between the brain *sans* imminent volitional action and the brain immediately before a volitional action.[22] Thus, even if the sequence of events prior to volitional action follows a circular pattern, there is still the need for a physical explanation for the mentioned difference. Unless interactive dualism is ruled out for independent reasons, the "ebb and flow" of brain activity could well be the effect, rather than the cause, of mental activity.

Therefore, the standard way of neuron firing brings us full circle back to the initial question: namely, what the physical explanation of the NCO is, if there is one at all. The following hypotheses are partly mechanisms taken from the neurophysiological literature (albeit not from a volitional actions context), and partly my own ideas based on standard physics and biophysics/biochemistry. Let me make their purpose clear: they serve as "placeholders" for the non-existing empirical research on the causal origin of volitional actions. I constructed them in order to give the physicalist a working hypothesis where dedicated empirical data are missing. To the extent to which those scenarios prove plausible, they buttress the physicalist's case. To the degree to which they prove implausible, interactive dualism becomes more viable as an explanation for the causal origin.

3.1 Causal Origin Outside the Body

A first natural suggestion is that the NCO is linked to a causal chain that leads out of the body. One such scenario is that of 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. For example, photons hitting our eyes could cause a conformational change in retinal proteins, which in turn cause bipolar cells to fire, which through a long chain or network of synaptic transmissions (as described above) running through the cortex to the muscles cause the arm to move. This would clearly be a perfectly acceptable, uncontroversially complete physical explanation.[23]

However, as indicated above, such stimulus-triggered actions take a different path through the brain than stimulus-free actions; by the same token, the studies underlying Haggard's overview are all "cue-free". Hence, this possibility is barred as long as neurophysiology does not come up with new discoveries linking volitional actions to external causes. However, this seems unlikely since, as indicated above, most experiments investigating

[22] For example, the Libet experiments showed that activity in the supplementary motor area (SMA) increases significantly and consistently shortly before the onset of muscle contraction.

[23] "Complete" at least in the sense that all involved physical events can in principle be explained by physics, even if for some a physical explanation has not yet been found.

volitional action rule out external causal influences. Therefore, we must make do with explanations inside the body.

3.2 Hormones Triggering the NCO

I suggest the hypothesis that the NCO is triggered by hormones coming from an endocrinal (hormonal) gland. An endocrinal gland does not consist of neurons, which is why it can only trigger, but not constitute, the NCO. It is well-known that hormones can act as neurotransmitters (in fact some substances function as both neurotransmitters and hormones, e.g. vasopressin and oxytocin). Let's call this the "endocrinal" hypothesis. The endocrinal hypothesis requires a brain area whose neurons are directly influenced by a hormonal gland.

I am, however, not aware of any brain areas along the abovementioned pathway that could fit the bill. The dopaminergic influence of the *substantia nigra* (SN) on the basal ganglia (cf. Haggard 2008, p. 936) unfortunately does not come into question as a candidate. 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) influences cortical processes (Pessiglione *et al.* 2006), 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 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 (Wilson 1999, pp. 191f, Hille 2001, Chap. 20).

3.3 Pacemaker Cells as NCO

The first two proposals have in common that they rely on the regular receptor-mediated generation of APs as described in section 2. But in principle APs can also be generated by a different mechanism, that of self-generation.

The neurons which are capable of self-generated APs are called *pacemaker cells*. Those neurons regularly self-generate APs through a cyclic mechanism of ion in- and outflux (Hille 2001, pp. 147ff). This makes them independent of external causes for AP generation. However, for all we know such pacemaker neurons exist in the heart only; it is doubtful that they can be found in the brain. Moreover, their activity is one of strict (though perhaps modifiable) regularity, which contradicts the idea of "irregularly willed" voluntary actions. Again, the possibility of identifying such neurons at the source of volitional actions can presently not be ruled out, but it is very unlikely.

3.4 Conformational Change of Neuronal Proteins as NCO

Neuronal function crucially depends on proteins such as the already mentioned Na^+/K^+ channels. Conformational change in such proteins is what drives their function (e.g. the transport of ions into or out of a cell), which in turn drives activities on the level of the whole cell (e.g. the issuing of an AP). It therefore stands to reason that conformational changes in proteins could in principle constitute the NCO. Consequently, the following hypotheses all involve conformational changes of some proteins. However, such change comes at a cost: it basically requires energy expenditure.[24]

At this point, some explanations about energy conservation are in order. If there is a complete physical explanation of volitional actions, then the energy expenditure needed for the conformational change of proteins will be accounted for in terms of the loss of energy of other physical systems – this is what energy conservation means. What we therefore need is a hypothesis that explains the energy change in proteins in terms of an energy *exchange* between those proteins and other nearby physical systems (e.g. ligand molecules). As the demand for a complete physical explanation dictates, we need other physical systems to account for the change in the proteins.

With this in mind, where might a conformational change in proteins take place such that this change is conducive to increasing a neuron's firing rate? One first option is that sodium or potassium channels open spontaneously or are caused to open in a deviating way by the binding of molecules, thereby triggering an AP.[25] Second, voltage-dependent Ca^{2+} (calcium) channels in the boutons might open without there having been a prior voltage change (i.e. without there having been an AP).[26] (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.[27] That would have the same effect as extracellular calcium flowing in. Fourth, neurotransmitter vesicles might spontaneously be released from the axon terminal by exocytosis,

[24] For a good overview of research concerning energy issues in neuronal protein conformational change see Wilson (1999).

[25] To be sure, there are so-called ligand-gated sodium channels whose occurrence is, however, restricted to the neuromuscular junction (*cf.* Hammond 2015, Chap. 6).

[26] Normally those channels respond only to large membrane depolarizations (Hammond 2015, p. 151).

[27] 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) (Hammond, p. 155). A release of calcium from there occurs normally upon an appropriate signal (e.g. the formation of inositol triphosphate) through Ca-permeable channels. The proteins primarily serve as calcium-binders to reduce cytosolic Ca^{2+} (which is toxic in too high concentrations) (Hammond, p. 51).

which also requires the conformational change of some proteins (Südhof 1995). The central question concerning all these options is how the respective proteins might be modified in ways that preserve energy. The following options come to mind:

1. Ersatz ligand molecules (distinct from neurotransmitters or hormones) 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,[28] but all of them are pharmacologically active chemicals supplied from outside. No endogenous substance is known to modify neuronal ion channels directly. There are still two protein targets for surrogate ligand molecules: the proteins involved in vesicular release from the bouton and the proteins involved in calcium release from intracellular buffers. As regards the former option, once again no endogenous substances come into question for such an effect; exogenous substances known to trigger vesicular release are classified as strong toxins.[29] The same holds true for calcium release from intracellular buffers (except that there are perhaps no toxins causing a "dam-breaking" of intracellular buffers).

Concerning (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 seems clearly too low to account for the dense volitional action patterns present in humans and, for that matter, animals;[30] but if it explains the baseline, it cannot also explain the *increase* in firing rate.

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

[29] 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 *et al.* 2008). This causes extreme pain and often death.

[30] 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} to 10^{-3} times per second and release site, which means it occurs once every 100 to 1000 seconds (Stevens 1993, p. 56).

Option (3): Could the NCO be triggered quantum-mechanically? It is important to distinguish this approach from the one proposed by Beck and Eccles (1992), Beck (1996, 2008) where the non-physical mind exploits quantum indeterminacies to trigger processes in neurons without violating energy conservation. Here, in exploring an energy-conserving, purely physical explanation of the NCO, we wish to know whether quantum mechanics alone, without the meddling of a non-physical mind, suffices to explain the behavior of neurons in volitional actions.

(1992; Beck (1996, 2008))

There are in general plenty of empirical problems with quantum-mechanical approaches to brain processes (for an overview see Cucu 2022, Chap. 6.3). Moreover, there is yet a further problem, namely that the frequency of quantum collapse events (independent of their vigor) does in all likelihood not match the occurrence pattern of volitional actions (Bourget 2004, Cucu 2020).

To fix the account, one might suggest a combination of "outlier molecules" and quantum events: quantum events explain the baseline and the "outlier molecules" the increase. The idea is *prima facie* tenable (provided the empirical problems with quantum events in the brain can be overcome). In fact, John Eccles, one of the most prominent defenders of a quantum-mechanical approach of neuronal firing, is prepared to relegate quantum events to explaining the baseline rather than the increase in firing rate, as Eccles (1994, Chaps. 4, 5) seems to suggest. However, it faces the intrinsic problems of the outlier hypothesis pointed out above. All in all, even a refined quantum account does not look as if it could satisfactorily explain the neuronal firing patterns in volitional actions.

In summary, the prospects of finding an energy-conserving, protein-based account of the NCO look bleak. One needs either special ligand molecules or a non-biochemical physical explanation to arrive at such an explanation. But neither seems to be available. Now, *if* the NCO consists in protein conformational change, and *if* indeed no physical explanation of that change can be found, then the conformational changes imply energy (and/or momentum) non-conservation. This is something that has been associated with the interaction of a non-physical mind (rightly so, as I will argue below).

Detractors of interactive dualism (e.g. Pollock 1989, p. 19, Flanagan 1991, p. 21, Westphal 2016, p. 42) typically use this to establish a reduction argument against mental interaction. The core of the argument is that, since conservation laws cannot be violated, interactive dualism must be false. The reasoning goes roughly like this:[31]

"against" in italics

[31] Or at least it should. Sometimes one finds the objection from energy conservation lumped together with something like a 'causal nexus objection", according to which the interaction of mind and matter does not take place because mind and matter are so radically different (thus in Dennett 1991, pp. 34f).

"

(P1) The total amount of energy (and momentum) in the universe is constant.

(P2) Interactive dualism entails that energy (or momentum) is added (or subtracted) from physical systems without being accounted for in the form of loss (or gain) in energy or momentum by other physical systems.

(C) Therefore, interactive dualism is false.

However, although (P2) is correct, (P1) rests on a misconception of conservation laws (see also Cucu and Pitts 2019, Pitts 2019). There are a number of problems with (P1). First, it is not even clear that the total amount of energy-momentum is well-defined (Peebles 1993, p. 139). If it is not well-defined, there is no way of telling whether it remains constant or not. Second, the mistaken version of energy conservation rests on the tacit assumption that the universe is a closed system (in which case, provided the pertinent quantities are well-defined, the amount of energy-momentum would indeed remain constant, per definition). However, such an assumption begs the question against interactive dualism (and divine interaction, for that matter), because if the universe were a closed system, no interaction would be possible, at least none that alters its energy or momentum content. To be sure, there are dualistic attempts to frame interactive dualism in a quantity-conserving manner (e.g. Broad 1937, Ducasse 1960, Hart *et al.* 1988, Lowe 1992); but they all fail (for an analysis of their problems see Cucu and Pitts 2019).

The third mistake is to conceive of conservation *globally* and *categorically.* Global conservation means that energy or momentum conservation is a property of the universe as a whole. However, conservation is *local* rather than global (applying to individual physical systems, not the universe as a whole). Also, conservation is not categorical (endowed with some sort of necessity, expressed by the word "cannot"); instead, it is *conditional.* This is how physicists understand conservation laws: as local and conditional. "Local" here means that conservation can be defined for each physical system individually, be it as small as an atom or as big as a supernova, while "conditional" means that it obeys an "if-then" pattern. More precisely, energy and momentum conservation are a matter of *bi*conditionality:

relevant Noether symmetry preserved $\equiv$ energy (or momentum) conserved

Could you replace this logical symbol (signifying a biconditional relation) by its alternative (↔)? I fear that some readers of M&M might not be familiar with logical notation. "↔" is in any event the more intuitive symbol.

The Noether symmetries are continuous (i) spatial and (ii) temporal symmetries.[32] "Continuous" means that the symmetry is preserved

[32] Cf. Noether 1918, Goldstein 1980, Chap. 12-7

through infinitesimally small alterations. For example, a sphere is continuously symmetrical with respect to rotation around its center, because no matter how small (or large) the angle of rotation, it will always be the same. A cube, by contrast, is not continuously symmetrical, because only rotations by 90 degrees yield symmetry. Spatial and temporal symmetry mean, simply put, that the physical system behaves the same no matter where in space or time it is located. An entailment of this is that if the relevant Noether symmetry is not preserved, then energy or momentum are not conserved. In other words, by the nature of its biconditionality,[33] the Noether theorem provides no categorical exclusion of non-conservation.

Now back to our question about protein conformation change being the NCO. According to the Noether theorem, energy and momentum non-conservation are not excluded *a priori.* All it says is that if the relevant symmetries are broken, then the corresponding quantity (energy or momentum) will not be conserved. In practice, however, it is extremely difficult, if not impossible, to ascertain whether the Noether symmetries are preserved or not.

A better way to check for conservation in open systems is the above-mentioned approach of comparing the in- and outflux of the quantity in question. The in- and outflux will add up to zero (over a period of time) only if the exchange occurs exclusively with adjacent physical systems; if a non-physical entity interacts, the sum of in- and outflux will not be zero.[34] Hence, an increase in energy unaccounted for by a decrease in neighboring systems is a sure sign of non-physical interaction. And that is exactly what seems to be happening with the hypothesized NCO, conceived of as a protein. I have considered a number of possible candidates for the energy increase required for protein conformational change, but none of them appears plausible. And, if no physical system can account for the energy increase required for protein conformational change, then the invocation of a *non*-physical entity is called for.

It is important to stress that this conclusion is not a (contingent) hypothesis offered to explain an otherwise unexplained phenomenon. Rather, it follows *ipso facto* from non-conservation (provided of course that the non-conservation is genuine, see above). Under the perspective offered here, non-conservation and the lack of a complete physical explanation become equivalent.

In this section, I have attempted different routes of coming up with

[33] A biconditional is true if either both its antecedent and consequent are either true or false, see standard logical truth tables.

[34] It should be noted that there are attempts at making interactive dualism quantity-conserving by ascribing the pertinent physical quantities to the mind (e.g. Collins 2011a). In that case the in- and outflux would of course cancel out. However, there are severe problems with such accounts, as pointed out by Cucu and Pitts (2019) and Cucu (2022, Chap. 5.4).

a complete physical explanation for volitional actions. The most plausible of them are still wildly implausible. Put differently, it seems as if positing energy non-conservation as part of the physical explanation is inevitable, which is equivalent to the failure of a complete physical explanation. And based on Noether's theorem, this in turn means that something non-physical must have interacted. Now, within the logical space spanned by the concept of "non-physical entity", the mind is the most immediate candidate.[35] After all, we know first-hand that we mentally cause volitional actions!

Of course, my argument so far is subject to the *caveat* that future research might discover hitherto unknown physical mechanisms which complete the picture so as to yield a complete physical explanation of volitional actions. Also, to repeat, to the extent to which non-physicalist ontologies other than interactive dualism provide a non-physical causal origin as explanation for volitional actions, they are *prima facie* as good alternatives to physicalism as interactive dualism.

At any rate, based on above arguments it seems that the physicalists' reliance on empirical neurophysiology, so cherished as a scientific foundation of physicalism, seems to undermine rather than support their position. At least for now, the tables are turned on physicalism.

4. Conclusion

According to physicalists, volitional actions must and do have a complete physical explanation. In contrast to many other areas of physiology, however, no detailed research into the mechanistic details of the neurophysiology of volitional actions has been undertaken. In any event, it is clear that volitional actions cannot be explained by physical factors outside the body; there is a known but different pathway for this that does not belong to volitional actions.

I therefore turned to available mechanisms for neuronal stimulation. Hormonal influence and pacemaker neurons have a different physiological role. Conformational changes of key proteins for neuronal excitation find no good underlying explanation either (outlier molecules and quantum events both are implausible for different reasons, as is a combination of them). Because of this, they entail a violation of energy conservation, which via the Noether theorem demands the interaction of something non-physical; I proposed this to be the non-physical mind. All in all, the picture gained by the present investigation looks much more like one in

[35] In principle, the non-physical entity in question could also be God, along the lines of divine occasionalism, where every physical event is directly brought about by God (and not by the powers of other physical systems). However, I shall not pursue this proposal further, if only because it raises the question of the existence of God.

which a non-physical mind gently interacts with the brain than one in which all brain events have only physical causes.

References

Akkal D., Dum R.P., and Strick P.L. (2007): Supplementary motor area and presupplementary motor area: Targets of basal ganglia and cerebellar output. *Journal of Neuroscience* **27**, 10659–10673.

Beck F. (1996): Can quantum processes control synaptic emission? *International Journal of Neural Systems* **7**(04), 343–353.

Beck F. (2008): Synaptic quantum tunnelling in brain activity. *NeuroQuantology* **6**(2).

Beck F. and Eccles J.C. (1992): Quantum aspects of brain activity and the role of consciousness. *Proceedings of the National Academy of Science of the USA* **89**, 11357–11361.

Bourget D. (2004): Quantum leaps in philosophy of mind: A critique of Stapp's theory. *Journal of Consciousness Studies* **11**(12), 17–42.

Brinkman C. and Porter R. (1979): Supplementary motor area in the monkey: Activity of neurons during performance of a learned motor task. *Journal of Neurophysiology* **42**(3), 681–709.

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

Brüntrup G. and Jaskolla L., eds. (2017): *Panpsychism: Contemporary Perspectives*, Oxford University Press, Oxford.

Chalmers D. and McQueen K. (2022): Consciousness and the collapse of the wave function. In *Quantum Mechanics and Consciousness*, ed. by S. Gao, Oxford University Press, Oxford, pp. 11–63.

Collins R. (2008): Modern physics and the energy-conservation objection to mind-body dualism. *American Philosophical Quarterly* **45**(1), 31–42.

Collins R. (2011a): A scientific case for the soul. In *The Soul Hypothesis*, ed. by M. Baker and S. Goetz, Bloomsbury Publishing, London, pp. 222–246.

Collins R. (2011b.): The energy of the soul. In *The Soul Hypothesis*, ed. by M. Baker and S. Goetz, Bloomsbury Publishing, London, pp. 123–133.

Cucu A.C. (2022): Interacting minds in the physical world. Dissertation, University of Lausanne, Lausanne, Switzerland. Accessible at `philarchive.org/rec/CUCIMI`.

Cucu A.C. and Pitts B.J. (2019): How dualists should (not) respond to the objection from energy conservation. *Mind and Matter* **17**(1), 95–121.

Cucu A.C. (2020): Does consciousness-collapse quantum mechanics facilitate dualistic mental causation? *Journal of Cognitive Science* **21**(3), 429–473.

Deecke L. and Kornhuber H.H. (1978): An electrical sign of participation of the mesial "supplementary" motor cortex in human voluntary finger movement. *Brain Research* **159**(2), 473–476.

Dennett D.C. (1991): *Consciousness Explained*, Penguin Books, London.

Ducasse C. (1960): In defense of dualism. In *Dimensions of Mind: A Symposium*, ed. by S. Hook, New York University Press, New York, pp. 85–90.

Eccles J.C. (1994): *How the Self Controls Its Brain*, Springer, Berlin.

Farris J.R. (2016): *The Soul of Theological Anthropology: A Cartesian Exploration*, Taylor & Francis, Oxfordshire.

Farris J.R. (2023): *The Creation of Self: A Case for the Soul*, John Hunt Publishing, Winchester.

Flanagan O.J. (1991): *The Science of the Mind*, MIT Press, Cambridge.

Fodor J. A. (1983): Making mind matter more. In *Mental Causation*, ed. by J. Heil and A.R. Mele, Clarendon, Oxford.

Fried I., Mukamel R., and Kreiman G. (2011): Internally generated preactivation of single neurons in human medial frontal cortex predicts volition. *Neuron* **69**, 548–562.

Fried I., Haggard P., Biyu J.H., and Schurger A. (2017): Volition and action in the human brain: Processes, pathologies, and reasons. *Journal of Neuroscience* **37**(45), 10842–10847.

Goldstein H. (1980): *Classical Mechanics*, Addison-Wesley, Boston.

Haggard P. (2008): Human volition: Towards a neuroscience of will. *Nature Reviews Neuroscience* **9**(12), 934–946.

Hammond C. (2015): *Cellular and Molecular Neurophysiology*, Academic Press, Boston.

Hart W.D. (1988): *The Engines of the Soul*, Cambridge University Press, New York.

Hasker W. (1999): *The Emergent Self*, Cornell University Press, Ithaca.

Hille B. (2001): *Ion Channels of Excitable Membranes*, Sinauer Associates, Sunderland.

Jahanshahi M.I., Jenkins H., Brown R.G., Marsden C.D., Passingham R.E., and Brooks D.J. (1995): Self-initiated versus externally triggered movements. 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), 913–933.

Kim J. (1998): *Mind in a Physical World: An Essay on the Mind-Body Problem and Mental Causation*, MIT Press, Cambridge.

Libet B., Gleason C.A., Wright E.W., and Pearl D.K. (1983a): 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**(Pt3), 623–642.

Libet B., Wright E., and Gleason C. (1983b): Preparation – or intention-to-act, in relation to pre-event potentials recorded at the vertex. *Electroencephalography and Clinical Nerophysiology* **56** 367–372.

Lowe J. (1992): The problem of psychophysical causation. *Australasian Journal of Philosophy* **70**(3), 263–276.

Lowe J. (2006): Non-Cartesian substance dualism and the problem of mental causation. *Erkenntnis* **65**, 5–23.

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

Meixner U. (2008): New perspectives for a dualistic conception of mental causation. *Journal of Consciousness Studies* **15**(1), 17–38.

Meixner U. (2019): Elements of a theory of nonphysical agents in the physical world. *Organon F* **26**(1), 104–121.

Montero B.G., and Papineau D. (2016): Naturalism and physicalism. In *Blackwell Companion to Naturalism*, ed. by K.J. Clark, Wiley, Hoboken, pp. 182–195.

Moreland J.P. (2018): In defense of a Thomistic-like dualism. In *Blackwell Companion to Substance Dualism*, ed. by J. Loose and A.J.L. Menuge, Wiley, Hoboken, pp. 102–122.

Moreland J.P., and Scott B.R. (2000): *Body & Soul – Human Nature & the Crisis in Ethics*, InterVarsity Press, Downers Grove.

Noether E. (1918): Invariante Variationsprobleme. *Nachrichten von der Gesellschaft der Wissenschaften zu Göttingen, Mathematisch-Physikalische Klasse*, pp. 235–257.

Owen M. (2020): Aristotelian causation and neural correlates of consciousness. *Topoi* **39**(5), 1113–1124.

Papineau D. (2001): The rise of physicalism. In *Physicalism and Its Discontents*, ed. by C. Gillett and B. Loewer, Cambridge University Press, Cambridge, pp. 3–36.

Peebles P.J.E. (1993): *Principles of Physical Cosmology*, Princeton University Press, Princeton.

Pessiglione M., Seymour B., Flandin G., Dolan R.J., and Frith C.D. (2006): Dopamine-dependent prediction errors underpin reward-seeking behaviour in humans. *Nature* **442**, 1042–1045.

Picard N., and Strick P.L. (1996): Motor areas of the medial wall: A review of their location and functional activation. *Cerebral Cortex*, **6**(3), 342–353.

Pitts, J.B. (2019): Conservation laws and the philosophy of mind: Opening the black box, finding a mirror. *Philosophia* **48**(2), 673–707.

Pollock J.L. (1989): *How to Build a Person: A Prolegomenon*, MIT Press, Cambridge.

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

Rickabaugh B. and Buras T. (2017): The argument from reason, and mental causal drainage: A reply to Peter van Inwagen. *Philosophia Christi* **19**(2), 381–399.

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), 744–754.

Roland P.E., Larsen N., Lassen A., and Skinhøj E. (1980): Supplementary motor area and other cortical areas in organization of voluntary movements in an. *Journal of Neurophysiology* **43**(1), 118–36.

Schurger A., Sitt J.D., and Dehaene S. (2012): An accumulator model for spontaneous neural activity prior to self-initiated movement. *Proceedings of the National Academy of Sciences of the USA* **109**(42), 2904–2913.

Soon C.S., Brass M., Heinze H.J., and Haynes J.D. (2008): Unconscious determinants of free decisions in the human brain. *Nature Neuroscience* **11**(5), 543–545.

Stein P.S.G. (1978): Motor systems, with specific reference to the control of locomotion. *Annual Review of Neuroscience* **1**(1), 61–81.

Stevens C.F. (1993): Quantal release of neurotransmitter and long-term potentiation. *Cell* **72** Supplement, 55–63.

Stoljar D. (2021): Physicalism. In *Stanford Encyclopedia of Philosophy Archive*, ed. by E.N. Zalta and U. Nodelman. Accessible at `https://plato.stanford.edu/entries/physicalism/`.

Südhof T.C. (1995): The synaptic vesicle cycle: A cascade of protein-protein interactions. *Nature* **375**, 645–653.

Swinburne R. (1997): *The Evolution of the Soul*, Oxford University Press, Oxford.

Swinburne R. (2013): *Mind, Brain, and Free Will*, Oxford University Press, Oxford.

Swinburne R. (2019): The implausibility of the causal closure of the physical. *Organon F* **26**(1), 25–39.

Taliaferro C. (1994): *Consciousness and the Mind of God*, Cambridge University Press, Cambridge.

Tanji J., and Kurata K. (1982): Comparison of movement-related activity in two cortical motor areas of primates. *Journal of Neurophysiology* **48**(3), 633–653.

Tanji J. and Keisetsu S. (1994): Role for supplementary motor area cells in planning several movements ahead. *Nature* **371**, 413–416.

Ushkaryov Y.A., Rohou A., and Sugita S. (2008): α-latrotoxin and its receptors. In *Pharmacology of Neurotransmitter Release. Handbook of Experimental Pharmacology, Vol. 184*, ed. by T.C. Südhof and K. Starke, Springer, Berlin, pp. 171–206.

von Wachter D. (2019): The principle of the causal openness of the physical. *Organon F* **26**(1), 40–61.

Weir R.S. (2023): *The Mind-Body Problem and Metaphysics: An Argument from Consciousness to Mental Substance*, Taylor & Francis, Oxfordshire.

Westphal J. (2016): *The Mind-Body Problem*, MIT Press, Cambridge.

Wilson D. (1999): Mind-brain interaction and violation of physical laws. *Journal of Consciousness Studies* **6**(8–9), 185–200.

Received: 24 October 2023
Revised: 14 December 2023
Accepted: 21 December 2023
Reviewed by two anonymous referees