Thinking Out of the Physicalist Box: A Surprisingly Simple Solution of the Free Will Dilemma

Alin Christoph Cucu[1]

Abstract

The dilemma of free will is that if volitional actions are caused deterministically, they are not free, and if they are caused indeterministically, they are not free either, because then they happen by chance and are not up to the agent. I assume that all volitional actions are caused by the brain. By a free action I understand an action which meets both of the following criteria: (i) the agent could have done differently than perform the action; (ii) the agent is the ultimate author of the action. I claim that the above dilemma is an entailment of physicalism because on physicalism, the agent must consist of or supervene on brain matter. The only processes available in the physicalist brain are deterministic and indeterministic physical processes, neither of which can constitute freedom in the relevant sense. It does not make any difference whether the physicalism in question is supervenient or reductive, because not even supervenience (on physical events) grants mental events enough ontological independence to meet criterion (ii). I conclude that purely physical systems cannot have freedom in the above sense. My claim is that a human being is essentially a soul (a non-physical substance) and has as an accidental part a body (a physical system), with the soul being able to causally influence the brain at will, thus meeting both criteria of free actions and constituting a third way between the horns of the dilemma. However, my solution faces the physicalist objection that such interactionist dualism entails violation of the principle of energy conservation (PEC); I will show that with an adequate version of PEC, this is not the case and that therefore interactionist dualism is a simple, powerful and robust solution to the free will dilemma. Finally, to undergird the validity of my solution, I will show how it helps close the gap in four contemporary libertarian accounts of free will.

I The Dilemma

The dilemma of free will is that "if volitional actions are caused deterministically, they are not free, and if they are caused indeterministically, they are not free either, because then they happen by chance and are not up to the agent." (Von Wachter 2003, 1)

Let me clarify the relevant terms first.
By *deterministic causation* I understand causation which is such that the effect-event was necessitated by antecedent events and conditions together with the laws of nature. Take gravitation as an example: If there are two planets with masses m_1 and m_2 (antecedent events) in a universe containing only those two planets (antecedent conditions) and the law of gravitation is in place, the two planets will *necessarily* be attracted by the force $F_G = G\frac{m_1 m_2}{r^2}$. This kind of causation is certainly part of, but to be distinguished from, the global doctrine of determinism (cf. Hoefer 2016).
By *indeterministic causation* I understand causation which is such that the effect-event is brought about with a certain (objective) probability by antecedent events and conditions together with the laws of nature. Consider a radioactive atom (e.g. Radium). There is a probability of 0.5 that a

[1] Internationale Akademie für Philosophie im Fürstentum Liechtenstein, FL-9493 Mauren; contact: acucu@iap.li

certain Radium atom will decay within the half-life of Radium (1600 years). Notice the contrast to the above deterministic example: The Radium atom does *not necessarily* decay within 1600 years; there is just a fifty-fifty chance that it does.

A *volitional action* is one done at will. By 'at will' I do not mean that the action is necessarily free (in the sense below), just that it seems to us from our introspection that it was our will that initiated the action. It is those actions and only those to which the free will dilemma applies; reflexes and actions induced by external factors like drugs etc. do not count as free from the outset. I share the common assumption that volitional actions are caused by the brain, most notably by the cortex; of course, this still leaves open the question what causes the brain to cause those actions, but this is in fact the main issue of this essay and I shall come back to it in the next sections. All *free actions* (if they exist) are necessarily volitional actions. However, it is not clear that the reverse is true, namely that all volitional actions are free; after all, it could be that our impression of free volition is just an illusion. To be a free action in my sense, a volitional action must meet both of the following criteria:

(i) The agent could have done differently than perform the action.
(ii) The agent is the ultimate author of the action.

Criterion (i) is important for free will because it maintains that a person could do a different action or refrain from doing the action. In other words, if the person could neither perform another action nor refrain from doing any action at all, then the action is forced by some factor outside the agent's control. This idea is especially relevant if one ties free will to moral responsibility (as is usually done). Suppose Albert (our agent) walks down the street and suddenly becomes aware of a very expensive mountain bike leaning against a lamp post without being locked. Suppose further that there is no one around to watch (not even a hidden camera). Now if Albert steals the bike (action A), it seems he is doing something wrong for which he is culpable. But what if Albert claimed that he just *had* to do A, so that the alternatives (leaving the bike in its place = action B; taking the bike to the municipal repository = action C) were completely unavailable for him? It seems that our blaming Albert depends on B and/or C being somehow (even if very weakly) available to him. If indeed it turned out that Albert was somehow *determined* to do A – say by a certain psychological dysfunction – so that he *could not possibly* do B (= refrain from A) or C (= do another action than A), we would probably not hold him accountable, because his action would not have been free. I conclude that "could have done otherwise" is a crucial hallmark of free actions. However, as we shall see now, it is a necessary yet not sufficient condition.

It is in fact criterion (ii) which must come alongside (i) to make an action truly free. Consider Albert once again. If it turned out that he could have done otherwise than steal the bike – e.g. because of a psychological dysfunction that makes him steal lonely bikes in 50 % of the cases – we will still not be inclined to call his action 'free'. It would be just a matter of the objective probability P = 0.5 that in one case he steals the bike and that in another case he leaves it untouched! My view therefore is (and that seems to me to lie at the bottom of our intuitions concerning free will) that Albert must be the *ultimate author* of this action in order for the action to be free. For an agent to be the ultimate author of his action, no prior cause can have determined the agent to will the action. This secures the agent's independence in initiating his action; if any prior factor determined the agent to will and thus initiate the action, this factor would take the role of ultimate 'authorship', precluding the agent from being the ultimate author of his actions. To be sure: on this picture, reasons can (and should) play an important role in the action-forming process, but they certainly cannot be (deterministic) *causes* for the action (cf. Pink 2011).

With these explications in place, it should become clearer why there is a free will dilemma at all. If (as many philosophers seem to assume) there are just the two types of causation mentioned – deterministic and indeterministic – actions must be constituted solely by one or both of them. If actions are caused deterministically, there is clearly no freedom, because criterion (i) is not met. If, on the other hand, actions are caused indeterministically, criterion (ii) is not met. A mix of both causation types, for example a causal chain some of whose parts are deterministic and others indeterministic, will not do either; no link of the chain would meet the criteria of freedom, and so the whole chain cannot be considered free. What is missing is a third kind of causation, on that meets both criteria. Traditionally, this third kind has been agent causation[2] in connection with an identification of the agent as an immaterial substance. But nowadays, most philosophers seem to follow Peter Strawson in considering agent (substance) causation as "obscure and panicky metaphysics" (Strawson 1962, 24). As I shall show in section II, however, it is precisely the move away from an immaterial agent towards a physicalist ontology that makes free will an obscure notion[3].

One last clarification before I move on. Of course, one could try to solve the free will dilemma by taking a compatibilist route. However, as compatibilism takes determinism to be true, the notion of free will advocated by compatibilists differs considerably from my definition, most notably in that it fails to meet criterion (ii). I therefore reject compatibilist approaches. My own position is libertarian; by the same token, I seek to mend only libertarian accounts with my solution of the free will dilemma (see section IV).

II It Is Physicalism that Creates the Dilemma

Where does the limitation to deterministic and indeterministic causation come from? It seems to me that it is commitment to physicalism that leaves one with just this limited scope, thereby creating the free will dilemma in the first place. Of course I cannot prove that every philosopher who sees the free will dilemma as a problem is a physicalist, nor do I wish to. What I want to do in this section is show that the free will dilemma is an entailment of physicalism. I take it that with respect to the problems outlined below, there are no relevant differences between physicalism, materialism (which are anyway often used interchangeably) and naturalism.

I will take *physicalism* to be the doctrine that everything that exists is physical or supervenes on something physical. Now if the agent is considered as just *being* some physical thing – presumably the brain – it is easy to see how this leads to the free will dilemma. A purely physical agent cannot be governed by other than physical processes, which in turn can be either deterministic or indeterministic; but, as I have shown in section I, neither deterministic nor indeterministic causation nor a combination of the two is a sufficient basis for freedom.

But there is another variety of physicalism, *supervenient* physicalism. Supervenient physicalism claims that non-physical properties supervene on physical properties; hence, mental properties (and the agent consisting of or bearing those properties) *supervene* on the brain matter. A set of properties A supervenes upon another set B "just in case no two things can differ with respect to A-properties without also differing with respect to their B-properties" (McLaughlin and Bennett 2018, 1). In other words, if two things have the same B-properties, they necessarily have the same A-properties. Hence, if two human beings have the same brain (B-)properties, they necessarily have the same mental (A-)properties. The mental properties should be understood as *global* properties of the brain and not as properties of the atoms or neurons constituting the brain (cf. (Lewis 1986, 14). Note also that physicalism thus formulated is a thesis with modal force, although its truth is normally considered to be contingent (Stoljar 2017, 9).

---

[2] Agent causation might be understood deterministically in the sense that whenever the agent wills an action, the action (or at least the brain process leading to the action) cannot fail to occur. But that is not the normal use of 'determinism', which refers to event causation only.

[3] Consider, for example, Peter van Inwagen's sobering conclusion that "free will is a mystery" (Van Inwagen 2000)

There are, however, strong objections against supervenient physicalism. Supervenient physicalism entails that two worlds with exactly the same physical properties must contain exactly the same mental properties. Against this, David Chalmers launched his argument of the conceivability of zombies (Chalmers 1996, 84-88). If it is conceivable that a world physically exactly like ours exists but in which human beings lack a conscious life, and if this conceivability entails metaphysical possibility (see Chalmers 2002), then supervenient physicalism is considerably weakened. Conversely, Descartes-type arguments like Richard Swinburne's (Swinburne 1996, Swinburne 2013, ch. 6) argue that it is conceivable that my body ceases to exist yet that it is inconceivable that I cease to exist; hence, my body and I (my mind) must be different things, again weakening the physicalist claim that there cannot be a mental life without a brain. Of course, none of those arguments is a final disproof of physicalism; they just pave the way for the plausibility of substance dualism.

But let us grant *arguendo* that supervenient physicalism is a viable position. One important stipulation that needs to be made in order to get physicalism off the ground with respect to freedom-conferring mental causation is that supervenient properties should be understood as "over and above" physical properties, i.e. as a *separate ontological category*. (The other option would be to consider them nothing "over and above" physical properties, that is, *identical* to physical properties, which makes freedom impossible, as I argued above.)
With this stipulation in place, does supervenient physicalism provide the necessary foundation for free actions? One indispensable requirement is that the supervenient mental properties need to have 'top-down' causal powers *of their own* (i.e. not be 'made to cause' by physical properties). But that already seems to go beyond the purview of supervenience. Consider again above example of the dotted-pattern-picture: no one in his right mind will claim that the supervenient property (the overall look of the picture) can by itself cause any change (of color or position) in any of its underlying dots. Or take 'supervenient sciences' as another, more elaborate example. Even if biological and chemical properties supervene on physical properties (like mass, charge, location), it seems wrong to claim that, say, a biological property like being a tissue changes the location of any of the underlying atoms, not to speak of a change in the charge of the electrons or the mass of the protons in that tissue.

At this point, it must be added that there is another relation apart from supervenience which might provide the required top-down causal powers: emergence (cf. Mumford and Anjum forthcoming). Roughly, emergence is the idea that parts of a whole give rise to novel properties in the whole, without those properties being reducible to the properties of the parts. I do not have the space here to deal with the vast literature on that subject; instead, I want to examine, by means of a neurobiological case study, what role an emergent agent/mind would play and how this squares with physicalism.

Case Study: Physicalism, Freedom and Neurotransmitter Release

Suppose, as neuroscientist John Eccles does (Eccles 1994), that volitional actions are triggered by the activity of neurons in the Supplementary Motor Area (SMA). Suppose further that this activity consists in neurotransmitter-filled vesicles (V) being released from a bouton of one of the neurons. If this is a process describable by classical physics, it definitely requires expenditure of energy. Where is this energy to come from? It might come through electric current traveling down the axon. But of course that just shifts the problem to the energetic trigger of the electric current; and, if in the causal history of the vesicle release only deterministic and/or indeterministic processes figure, the triggering of the action was not due to a free choice of the agent. Alternatively, the energy might come from other particles or enzymes floating around, who have the required kinetic and/or chemical energy. But the movements of such particles are obey the laws of statistical thermodynamics. It then seems that the vesicle release and the resulting

macroscopic behavior are in a relevant sense random and thus unfree. What would the emergent agent-mind have to do in order to initiate a free action? It seems that if action-triggering brain processes obey the laws of classical physics, the emergent agent-mind would either have to create additional energy *ex nihilo* or somehow *direct* the relevant molecules to their target location. Both options are clearly at odds with the laws of physics and therefore with physicalism. It seems, therefore, that an emergent mind providing free will must contradict physicalism, at least if the underlying physical picture is classical.

There is, however, the possibility that action-triggering processes are quantum-mechanical in nature. This means, roughly, that with the same initial conditions (energy, mass, location…), different outcomes are possible! For example, a particle small enough to be governed by quantum laws might change its position without there being any additional energy input, thus triggering the synaptic release. John Eccles and Friedrich Beck calculate that a particle as heavy as six hydrogen atoms might still obey the "quantal regime" (see Beck and Eccles 1992), which makes look the quantum mechanical approach look quite plausible. However, a different quantitative approach by David Wilson (Wilson 1999) yields the result that quantum processes cannot have any significant impact on neurons. Further, quantum processes obey statistical laws with objective probabilities. For example, an electron in the spin superposition state $|\uparrow>+|\downarrow>$ can, upon measurement, either exhibit the state $|\uparrow>$ or $|\downarrow>$ with equal probabilities of 0.5 each. It seems that if quantum processes are responsible for action-triggering, they do so with objective probability, thereby precluding freedom (see section I); alternatively, the agent's choice might override the objective quantum probabilities, which seems impossible on physicalism. Again, it looks that an emergent agent either complies to the laws of physics but cannot account for freedom, or must be something non-physical.

In summary, even an emergent account of the agent, if it is to fit the physicalist picture, cannot account for freedom of action. But of course, the emergent mind's novelty might just consist in the mind being non-physical and therefore not subject to laws of nature but being able to willfully influence physical processes. In fact, substance dualist accounts of emergent souls are being defended (Hasker 1999, Lowe 2006, Meixner 2004). The substance dualist account I am now going to present is leaves it open whether the soul is emergent or not, but nothing turns on that. My claim, and the main point of this paper, is that it can solve the free will dilemma.

III Interactionist Substance Dualism as the Solution

Substance dualism (SD) claims that human beings consist of two substances: the body, a physical substance, and the mind or soul, a mental (non-physical) substance (whether it should be construed as purely mental or partly physical will be considered below). A human agent *is essentially* her soul; she *accidentally has* a body. At this point, it is helpful to make more explicit what I mean by 'physical' and 'mental', respectively. Following Richard Swinburne (Swinburne 2013, 67-68), I define a physical property as one to which public access is possible, i.e. which can, at least in principle, be observed by anyone. Conversely, by a mental property I mean a property to which only its bearer has private access.

Interactionist substance dualism (ISD) claims that the soul can cause brain events, which in turn can cause bodily movements. Free action can be represented as follows on an interactionist substance dualist account:

1) The soul, that is the human agent, wills an action[4]. There is no prior cause to this willing (or volition) in the deterministic or indeterministic sense; reasons may of course play a role in leading to the willing, but they do not cause it either deterministically or indeterministically; the agent just "wills the willing".

[4] My terms 'willing'/'volition' are synonymous to Richard Swinburne's 'tryings' or Daniel von Wachter's 'choice events'.

2) Normally, the willing leads to a brain event, which then, if the neural machinery works properly, deterministically leads to a bodily movement.

On this account, both criteria for free action are met. The agent might will a different action or choose to not will any action at all (criterion i) of "could have done otherwise"); also, the agent is the ultimate source of her actions, because there is no prior cause to the willing (criterion ii)). To be sure, this does not entail that all of the agent's actions are free, nor does it entail that the agent keeps her ability to 'will her will' forever. It does entail, however, that under normal circumstances at least some of an agent's actions are genuinely free. This is what, to my mind, a libertarian account of free will should provide.

The PEC objection to ISD answered

ISD takes the mind or soul to be at least partly non-physical. One prominent objection that arises from this assumption is the so-called objection from the principle of energy conservation (PEC). It roughly states that the mind's interaction with the brain requires that the brain's total energy increases, which is taken to be a violation of PEC. I deal with this objection specifically, first because it is so widespread, and second because it bears direct relevance for freedom. After all, any manipulation of a physical system seems to require expenditure of energy; alternatively, if there are processes involving physical entities which don't require expenditure of energy, ISD might get a free pass with respect to PEC.

Let us first see what the PEC objection consists in. In a very coarse-grained analysis, Daniel Dennett (1991, 35) writes:

> 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. (Brackets added)

Implicit in Dennett's argument seems to be a version of PEC that we were presumably all taught in school. Following Robin Collins (2008), I call it CPEC (closed version of PEC):

> (CPEC) The amount of energy in the physical universe remains constant.

From there, the PEC objection seems to run as follows (let α be the mind and β be the brain):

1) The amount of energy in the physical universe remains constant. (CPEC)
2) According to ISD, α causally acting on β leads to an increase of energy in β.
3) α is purely non-physical and hence not part of the physical universe.
4) Therefore, ISD violates CPEC. (from 1, 2, 3)

The argument is sound, but the problem with CPEC is that it precludes ISD *a priori*, because it posits the universe as an isolated system. 'Isolated system' is a term from thermodynamics and denotes a system that does not exchange either matter or energy with its environment. If this applies to the universe, then *a fortiori* it applies to any brain within the universe, precluding any energetic influence of an immaterial mind. But as some philosophers have pointed out (Larmer 2014; Plantinga 2007, 126), positing the universe to be closed in the above sense is a metaphysical claim and goes beyond what can justifiably be called a scientific principle. After all, we – and physics, in particular – do not know whether no energy and/or matter can be added to (or subtracted from) the universe!

Because of that, Robin Collins offers what he calls the *boundary version* of PEC (in short BPEC):

> According to BPEC, the rate of change of total energy (or, more accurately, stress-energy) in a closed region of space is equal to the total rate of energy (or, more accurately, stress-energy) flowing through the spatial boundary of the region. (Collins 2008, 34)

BPEC avoids the abovementioned drawbacks of CPEC, but still faces a problem: It doesn't seem to make sense to talk of the mind, itself assumed to have no spatial location, sending energy *across the brain's boundaries*. It seems we need to adjust BPEC a little:

> (BPEC*) The rate of change of total energy in a closed region of space is equal to the total rate of energy flowing through the spatial boundary of the region plus the total amount of energy brought about in it non-locally.

Is it legitimate to speak of non-locally caused energy changes? Robin Collins thinks it is, because there seems to be at least one physical case in which energy increase is brought about non-locally, namely the increase of gravitational energy in General Relativity (GR). Collins writes that "no local concept of stress-energy (and hence energy-momentum) can be defined for the gravitational field in GR." (2008, 36) In other words, the gravitational field brings about an energy increase in a spatial region, even if the body which exerts gravitational force on the region is very far away; additionally, no flow of energy across the region's boundaries can be defined, although there is a clear energy increase within the region. He concludes that although BPEC is, strictly speaking, not violated in those GR cases, it simply doesn't apply to them. I therefore take BPEC* to be a version of PEC that both satisfies physicists' needs but leaves room for ISD.

At this point, it is tempting for a substance dualist to lean back and be content with BPEC* as a version of PEC which at least does not preclude ISD *a priori*. However, the mere fact that there is a physical process out there in which energy flow cannot be defined does not entail that ISD works without the violation of energy conservation. After all, is it not the classical picture of substance dualism that the soul is a purely mental (i.e. purely non-physical) substance? It seems, however, that even if energy cannot be *defined* for the mind, it would still have to *bear* energy. Retaining the classical picture, on the other hand, leads one to abandon the above GR-analogical interaction model. To sort out the options available to the substance dualist:

a) Assume a partly physical mind which exchanges energy with the brain (whether or not that energy can be defined).
b) Assume a completely non-physical mind. The energy needed for mind-brain-interaction is created *ex nihilo* by the mind.
c) Be agnostic about the ontological nature of the mind. Claim that the mind exploits quantum indeterminacies for mind-brain-interaction.
d) Be agnostic about the ontological nature of the mind. Claim that mind-brain-interaction does not require energy expenditure at all.

I cannot exhaustively deal with all the intricacies of the outlined strategies. In the remainder of this section, I would just like to present a survey of extant accounts falling under the headlines a) to d) and briefly comment on them.

Along the lines of a) are Robin Collins's abovementioned GR analogy model and his 'guitar-string model' of the soul (Collins 2011). To the former, Brian Pitts objects that GR makes ISD *harder*, not easier, especially on atheism (Pitts 2018). The latter seems to me to be more promising, especially as it has the potential to be tested empirically (Collins 2011, 243-44).

Something like b) has, to my knowledge, not yet been defended. I dare say that it does not seem completely outlandish; after all, the metaphysical assumption that energy cannot be created

(implicit in CPEC) is a priori on a par with the assumption that it *can* be created, and if God exists and acts according to what the monotheistic religions claim, we might even have a precedence for the latter.

Quantum-mechanical accounts along the lines of c) have been adopted or at least considered by (among others) Richard Swinburne (2013, 112-117), John Eccles (1986, 1994), Henry Stapp (e.g. Stapp 2007), Hans Halvorson (2011) and Alvin Plantinga (Plantinga 2011; with respect to God's interaction with the world). I already canvassed the idea behind them in section II, so I will not do that again here. As attractive as quantum mechanics may seem for the dualist, two issues arise. The first is whether it would not constitute a violation of the probabilistic laws if the mind could manipulate them, which seems necessary for true freedom of will. A second issue is that such an approach finally does force the dualist to show his colors concerning the ontological nature of the mind, i.e. whether it is purely mental or partly physical. Maybe Halvorson (2011) is right and (pure) mental events are what accounts for quantum events; but that is a matter of dispute, and so the dualist arguing quantum-mechanically needs to consider the partial physicality of the mind as well.

Finally, strategies following the outline of d) seem to be particularly attractive for maintaining a scientific worldview. If no expenditure of energy is needed, then no problems arise with respect to the mind's (non-)physicality or such 'weird' views as energy creation *ex nihilo*! Robin Collins (Collins 2008, 38-39) proposes EPR phenomena as a possible precedence. In EPR phenomena, two (or more) particles (e.g. photons, electrons…) are entangled. That means that a change in the state of one particle – for example, transitioning from a superposition state into an eigenstate – correlates with an *instant* change of the other particle into the *same* state. This has been shown to occur no matter how far apart the particles are, such that if a causal interaction between the particles is assumed, it would have to be a superluminal one! More importantly, though, the entanglement correlation does not include any exchange of energy. According to Collins, substance dualism might make use of this physical precedence by claiming that the mind-brain-interaction is EPR-analogical.
Two issues arise here. The first pertains again to the mind's ontological nature. The particles in EPR phenomena are all physical, so it seems the mind acting EPR-analogically on the brain would have to be partly physical, or else the analogy loses much of its power. The other issue concerns the nature of the EPR phenomena. If considered causal, they violate a fundamental principle of Special Relativity, namely that no transmission faster than light is possible. If, however, they are considered mere non-causal correlations, one must wonder how such a process helps the substance dualist along. To be sure, non-causal mind-brain correlations have been defended (parallelism and Leibniz's pre-established harmony), but each of them is a considerable move away from the original project of ISD.

In summary, there are interesting options for ISD to be spelled out in a coherent and science-friendly way, but to adjudicate between them, more research needs to be done.

IV How Interactionist Substance Dualism Helps Close Libertarian Gaps

So far, I have argued that a libertarian account of free will requires the agent to be an immaterial substance which is able to influence brain processes at will. In this final section, I would like to corroborate this thesis by examining four exemplary contemporary libertarian theories. I will show that two of them fail to solve the free will dilemma altogether because they rely solely on physical processes, while the other two weaken their merits by refusing to identify the agent as an immaterial substance.

Laura Ekstrom's preference theory

The centerpiece of Laura Ekstrom's account is the notion of 'preference'. By preferences she understands desires that have passed a process of critical evaluation with one's conception of the good (Ekstrom 2011, 371). The formation of such a preference, she maintains, is an action. At the same time, she holds that preferences are brought about *indeterministically*. Now a preference thus acquired leads to a decision or performance of an action *deterministically*. She considers those actions to be free:

> The resulting decision output, the preference, when indeterministically caused and noncoercively formed, is authored by the agent, since it is formed by her for reasons that justify and explain it, and its claim to being authentic is not defeated by the objection that she formed it because she had to, because it was causally necessitated by the past and the natural laws. (ibid., 373)

The problem with this approach is that an action is declared free even if it is brought about by events none of which is a free action. Recall that neither deterministic nor indeterministic causation can account for freedom. According to Ekstrom, the preference is formed indeterministically (beyond the agent's control, meeting criterion (i) but failing to meet criterion (ii)). The upshot is that the agent seems to have no control over his actions!

To be sure, Ekstrom (ibid., 375-76) attempts to salvage her account by arguing that the chance element in the indeterministic causation of a preference should be understood in a probabilistic sense, i.e. that there was a probability $0 < p < 1$ of the event occurring and a probability $q = 1 - p$ of the event not occurring. Thus, she maintains, one need not understand chance as a weird force ("Chance") determining the outcome or as the mere absence of purposiveness. Ekstrom holds that her account yields exactly what is required for purposive agential control, namely considerations (indeterministically) entering the agent's mind and subsequent deliberation. As I understand Ekstrom, the chance would then consist in it not being determined which considerations enter the agent's mind. The subsequent deliberation cannot be chancy, as Ekstrom insinuates with the use of the term "purposive".

However, this explication does not fix the account. Granted, considerations popping up indeterministically do not destroy free will, as long as the agent has the ability to pick one as her preference. But at this point it becomes obvious what is lacking in Ekstrom's account. As described before, the views the preferences to come about indeterministically, which precludes what I called in section III 'willings'. In fact, to my knowledge Ekstrom rejects any ontology of the agent that could do the work of the account given in section III[5].

Ekstrom's account is a good example of event-causal accounts libertarian accounts[6]. They all rely on indeterminism in a causal chain of events. Randolph Clarke and Justin Capes (2017) nicely sum up the problems all event-causal accounts face due to this assumption:

> If this strategy [of relying on indeterminism to salvage alternate possibilities] succeeds in showing that the required indeterminism would not undermine responsibility, it leaves unaddressed the charge that the requirement is superfluous, that it secures nothing of value that could not exist in a deterministic world. And it is hard to see how this charge can be answered. (21)

I therefore take it that event-causal libertarian saccounts do not have the resources to account for free will because there is no substance-agent to pick between the options which rational deliberation, or indeterministic processes for that matter, offer to the agent's mind.

---

[5] For example, the entry "Incompatibilist (Nondeterministic) Theories of Free Will" in the *Stanford Encyclopedia of Philosophy* (Clarke and Capes 2017) renders Ekstrom holding that "the agent *is* her preferences" (12).

[6] In my view, such accounts should be reclassified as compatibilist.

Robert Kane's agent-causal account

Robert Kane defends an agent-causal account. That is, he views the agent as a substance, in contrast to event-causal accounts, according to which events are all there is to free action. Kane also makes use of indeterministic processes in his theory, but he takes them to be a hindrance rather than a positive causal contributor to deliberation and action (Kane 2011, 393). In Kane's picture, an agent faced with a decision runs parallel efforts to do both actions; indeterminacy in the brain processes is a hindrance to one option or both; the option effectively instantiated is the one whose hindrances the agent overcame.

What is not entirely clear from Kane's account is what ultimately settles between two or more options. He might be read as claiming that it is indeterminacy that hinders one option so strongly that the other one is materialized. But to me that sounds like a great loss of control over the outcome and in moral responsibility. Take the example Kane himself gives: a businesswoman walks to a meeting crucial for her career and suddenly sees the injured victim of an ambush lying in the bushes. Now according to Kane, two efforts run parallel in the businesswoman's mind, namely (i) trying to ignore the victim and get on to her meeting and (ii) trying to stop and help the victim on pain of missing her meeting. Which one she will do ultimately depends on which of the brain processes corresponding to (i) and (ii), respectively, will be less hindered by indeterminacy. Let's assume the businesswoman has recently become aware of her selfishness and desperately wants to improve her character. The victim offers her a chance to do so. On Kane's account, if she fails to help the victim, she might attribute this outcome to indeterministic brain processes; if she believes Kane's theory, she might say to herself, "I tried to help the victim, but indeterministic processes in my brain suppressed the altruistic option in such a way that the career-oriented action won over". Perhaps Kane would reply that if prior to the incident she had made a self-forming action (SFA) towards being more altruistic, the outcome would be different, because the prior probabilities would be much more in favor of the altruistic action. At this point – like with Ekstrom's 'preferences' – it becomes crucial how those SFAs come about. But more about that in a minute.

The alternative reading of Kane is that for the settling of an action (against indeterministic harassing fire) no indeterminacy plays any role, but the decision of the agent (presumably guided by reasons). If so, the question arises which ontological nature the agent has – the answer is relevant not to actions only, but also to SFAs (see above). Kane's view is that the agent is a persistent substance, but which can be reduced to "states of affairs, events and processes involving it" (ibid., 396). To him, agents are "systems (…) in which new emergent capacities arise as a result of greater complexity or as the result of movement away from thermodynamic equilibrium toward the edge of chaos." (ibid., 396).

Again, as in section II, I do not wish to dive into the debate about emergence. Clearly, Kane has a reductive version of the emerging agent in mind (which makes it doubtful whether talk of emergence is still legitimate). As I argued in section II, such an emergent agent-mind cannot contribute anything to freedom of action; it seems totally subject to physical processes. As Kane seems unwilling to posit a non-physical agent with the powers to willfully change the course of brain processes, his account cannot provide a basis for freedom of action.

Carl Ginet's non-causal account

Carl Ginet (e.g. Ginet 2007) holds that a free action is neither caused deterministically nor indeterministically, but uncaused. As this seems puzzling, Ginet makes his claim clearer by assuming that the "causally simplest events that count as actions are mental events, like decisions and volitions" (ibid., 244). These mental events, Ginet claims, are uncaused in free actions. He grants that the subsequent (perhaps bodily) action is in turn caused by those mental events, but maintains that the action as a whole is still uncaused, because its first stage triggering the subsequent stages was uncaused (ibid., 245). In order to distinguish his view from causal pictures, he states that "given that an action was uncaused, all its agent had to do to make it the case that she performed that action was to perform it" (ibid., 247). Whereas on a causal picture: "For any event e, S made it the case that e occurred only if S *caused* e to occur." (ibid., 246; emphasis added). Notice how

close Ginet's picture is to the one I painted in section III: Ginet takes mental events (especially volitions) as uncaused in the sense that they just occur upon the agent's performance of them – just like on ISD. To see how much the two accounts concur, let us look at how Ginet answers two objections against his theory (which might equally be raised against ISD).

The first objection is that if an action is uncaused, it must be a matter of chance or luck, thereby stripping the agent of all control and accountability. But Ginet correctly answers this by maintaining that the luck objection applies to indeterministic/probabilistic causation only; any probabilistic understanding of chance does not apply to uncaused events. Only if an action like telling the truth is indeterministically caused, it makes sense to speak of a probability applying to it. Let's assume that to action A a probability of $p = 0.57$ applies. If A is indeterministically caused, this means that on 100 instances in which prior events and circumstances conducive to A obtain, only in 57 of those 100 instances A occurs, this ratio being a law of nature. In other words, we could *predict* that of 100 instances 57 lead to the occurrence of A. If A is uncaused, however, we could not predict at all the outcomes of the 100 instances. Hence, it makes no sense to speak of a probability, if by that a law of nature is meant; we could just say *retrospectively* that of 100 instances, 57 led to the occurrence of A, not being justified in contending that the next 100 instances will exhibit the same ratio. Ginet thus makes use of the fundamental distinction between free actions and probabilistic events, which I also pointed out in section I.
The second objection takes up motives and reasons as necessary explanations for actions. It goes as follows (ibid., 251):

(a) an action could have been up to the agent only if it has an explanation in terms of the agent's motives or reasons; and
(b) its being uncaused would preclude such an explanation.

Ginet replies by arguing that explanations need not be causal explanations. Therefore, even if a reason is the explanation for an action, the reason need not have *caused* the action. The relation between reasons/motives and actions is then not a *causal* one, but rather an *intentional* one: "[M]y reason for doing A was that I wanted to obtain B and believed that by doing A I would obtain B." (ibid., 251) A similar view is proposed by Thomas Pink (Pink 2011). ISD likewise allows for reasons to play an important role in bringing about an action (e.g. an explanatory one), but rejects to assign a causal role to them (see section III).
Because of those similarities between Ginet's theory and ISD, I take it that his account has the potential to account adequately for free will. Its weakness, as I see it, lies rather in what it does not say than in what it says. Ginet speaks of the agent making it the case that he triggers an action by non-causally bringing about a volition. However, Ginet himself is not clear about whether he views the agent as a substance or not (Ginet 2007, 245). He maintains that the question of the agent's ontological status is irrelevant to the question whether an action can be uncaused and up to the agent. I hope to have made it clear though that, *contra* Ginet, it *is* relevant. As we have seen, the non-causal view is still committed to actions having an explanation. That means, proponents will reject actions as being due to *ontic chance*, i.e. having neither a cause nor an explanation. Now the agent figures prominently in Ginet's (and Pink's) theory; it is his reason-based intentionality that makes actions happen. To me, this cries out for an account of what sort this being - that can bring about volitions "ex nihilo" - is. I, for one, cannot see a better candidate than an immaterial soul.

Von Wachter's theory of agent causation

The libertarian account which comes closest to ISD is Daniel von Wachter's theory of agent causation (Von Wachter 2003, Von Wachter 2009). He takes agents to be capable of bringing about what has been labeled in various ways 'efforts' (Kane 2011), 'volitions' (Ginet 2007), 'endeavorings' (Chisholm 1976) or 'tryings' (Swinburne 2013); his own term is *choice events*. Von Wachter uses this term to distinguish those events (also called by him *initial events*, because the initiate an action) from

events that are part of physical processes. Importantly, according to von Wachter, choice events are not caused by prior events, but by the agent. He leaves open whether choice events are mental or physical events (presumably brain) events. Also, his view allows, but does not depend on, indeterminism. He rejects classical determinism altogether and instead proposes a "directedness" theory of causality (Von Wachter 2009; Von Wachter 2012) in which things together with states of affairs form a basis for a tendency towards another state of affairs that is "deterministically" realized unless another tendency incompatible with the first one prevents it. Thus, the choice event together with the brain matter form a basis B_1 for a tendency T_1 to action A (say raising one's arm), which happens unless another tendency T_2 incompatible with T_1 (e.g. a cramp in the arm muscle) prevents it from occurring. No indeterminism is required to leave open alternate possibilities, but of course indeterminism might be adduced along the lines of Plantinga's Divine Collapse Causation (Plantinga 2011) or Eccles's and Beck's theory (Beck and Eccles 1992).

On that picture, Von Wachter rightly claims to have solved the dilemma of free action:

> With choice events the dilemma of free action is solved. Free actions are neither caused deterministically nor are they uncaused or indeterministically caused. Choice events are the third way that avoids both horns of the dilemma. (Von Wachter 2003)

Again, his view is very close to ISD. Actions begin with a choice event brought about by the agent, in the way I described it in section III; those choice events are neither deterministic nor indeterministic, but constitute a third way through the horns of the dilemma. The bodily movement may or may not happen, no absolute determinism is assumed here (although his directedness theory may be seen as representing what I called 'deterministic causation' above). However, Von Wachter leaves it open whether the agent is a soul or a brain, and correspondingly whether choice events are mental or physical events. As I argued in section II, physical systems lack the requisite properties for freedom; thus, Von Wachter might complete his account by defending a substance dualist version of it.

V Summary

I claim to have shown that the free will dilemma arises necessarily when human agents are construed in the physicalist sense, irrespective of whether one prefers an emergent, supervenient (non-reductive) or reductive physicalism. This is because physical systems (with or without emerging or supervenient properties) lack the type of causation requisite for freedom in the libertarian sense. I also claim to have presented a coherent account which constitutes a third way through the horns of the dilemma, namely interactive substance dualism (ISD). I defended ISD against the prominent objection from the principle of energy conservation (PEC) and showed that there are different strategies open for the substance dualist to spell out his view in a science-friendly way. Finally, I examined several contemporary libertarian accounts and showed that they fare the better the closer they move towards ISD.

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.
Chalmers, David. 1996. *The Conscious Mind*. New York: Oxford University Press.
———. 2002. "Does Conceivability Entail Possibility?" In *Conceivability and Possibility*, edited by T. Gendler and J. Hawthorne, 145–200. Oxford University Press.
Chisholm, Roderick. 1976. "The Agent as Cause." In *Action Theory*, edited by M. Brand and D. Walton, 199–211. Dordrecht: Reidel.
Clarke, Randolph, and Justin Capes. 2017. "Incompatibilist (Nondeterministic) Theories of Free Will (PDF Version)." *Stanford Encyclopedia of Philosophy*. https://plato.stanford.edu/entries/incompatibilism-theories/.
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.
Dennett, Daniel C. 1991. *Consciousness Explained*. Penguin Books.
Eccles, John C. 1986. "Do Mental Events Cause Neural Events Analogously to the Probability Fields of Quantum Mechanics?" *Proceedings of the Royal Society* 227: 411–28.
———. 1994. *How the Self Controls Its Brain*. Springer.
Ekstrom, Laura. 2011. "Free Will Is Not A Mystery." In *The Oxford Handbook of Free Will*, 2nd ed. Oxford University Press.
Ginet, Carl. 2007. "An Action Can Be Both Uncaused and Up to the Agent." In *Intentionality, Deliberation, and Autonomy*, edited by Lumer, 243–255. Ashgate.
Halvorson, Hans. 2011. "The Measure of All Things: Quantum Mechanics and the Soul." In *The Soul Hypothesis*, edited by Marc Baker and Stewart Goetz. continuum.
Hasker, William. 1999. *The Emergent Self*. Ithaca, N.Y: Cornell University Press.
Hoefer, Carl. 2016. "Causal Determinism." *Stanford Encyclopedia of Philosophy*. https://plato.stanford.edu/entries/determinism-causal/.
Kane, Robert. 2011. "Rethinking Free Will: New Perspectives On An Ancient Problem." In *The Oxford Handbook Of Free Will*, edited by Robert Kane, 2nd ed. Oxford University Press.
Larmer, Robert. 2014. "Divine Intervention and the Conservation of Energy: A Reply to Evan Fales." *Nternational Journal for Philosophy of Religion* 75 (1): 27–38.
Lewis, David. 1986. *On the Plurality of Worlds*. Oxford: Blackwell.
Lowe, Jonathan. 2006. "Non-Cartesian Substance Dualism and the Problem of Mental Causation." *Erkenntnis* 65: 5–23.
McLaughlin, Brian, and Karen Bennett. 2018. "Supervenience." *Stanford Encyclopedia of Philosophy*.
Meixner, Uwe. 2004. *The Two Sides of Being: A Reassessment of Psycho- Physical Dualism*. Paderborn: Mentis.
Mumford, Stephen, and Rani Lill Anjum. forthcoming. "Emergence and Demergence." In *N.N.* Taylor & Francis.
Pink, Thomas. 2011. "Freedom And Action Without Causation: Noncausal Theories Of Freedom And Purposive Agency." In *The Oxford Handbook of Free Will*, edited by Robert Kane, 2nd ed. Oxford University Press.
Pitts, Brian. 2018. "General Relativity, Energy Conservation, and Mental Causation: Carroll's Foundling." presented at the Philosophy of Physics One Day Conference, University of Cambridge, May 30.
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.
———. 2011. *Where the Conflict Really Lies: Science, Religion, and Naturalism*. 1st ed. Oxford University Press.
Stapp, Henry P. 2007. "Quantum Mechanical Theories of Consciousness." In *The Blackwell Companion to Consciousness*, edited by M. Velmans and S. Schneider. Blackwell.
Stoljar, Daniel. 2017. "Physicalism." *Stanford Encyclopedia of Philosophy*.

Strawson, P. F. 1962. "Freedom and Resentment." *Proceedings of the British Academy* 48: 1–25.
Swinburne, Richard. 1996. "Dualism Intact." *Faith and Philosophy* 13 (1): 68–77.
———. 2013. *Mind, Brain, and Free Will*. Oxford: Oxford University Press.
Van Inwagen, Peter. 2000. "Free Will Remains a Mystery." *Philosophical Perspectives 14: Action and Freedom*, 1–19.
Von Wachter, Daniel. 2003. "Free Agents as Cause." In *On Human Persons*, edited by K. Petrus, 183–94. Frankfurt/Lancaster: Ontos Verlag.
———. 2009. *Die Kausale Struktur Der Welt*. Alber.
———. 2012. "Kein Gehirnereignis Kann Ein Späteres Festlegen." *Zeitschrift Für Philosophische Forschung* 66 (3): 393–408.
Wilson, David. 1999. "Mind–Brain Interaction and Violation of Physical Laws." *Journal of Consciousness Studies* 6 (8–9): 185–200.