[Work in progress. Draft of 17.11.2016]

The dogma of physical closure[1]

1 The aim of the paper

The thesis I am considering is that events or states of affairs that are not, or cannot be reduced to, physical events or states of affairs (if there are such events and states of affairs at all), make no difference to the course of physical events, or to the succession of the physical states of the universe.

If true, this thesis has an enormous bearing on the philosophy of mind, and on the philosophy of agency and freedom. Yet, philosophers who use it as a premise to their positions rarely stop to make a serious attempt to convince their readers about its truth. Nor is it very common even to refer the reader to a work which is supposed to have established it authoritatively. This sort of treatment could suggest that physical closure is a well-established empirical fact, a datum that scientifically informed philosophers would accommodate in their theories without feeling much need to go over the justification for it, which has already been dealt with, and is perhaps mostly the business of the physics department and other science departments anyway. But it may also be a symptom of a large part of the philosophical community having grown accustomed to take it for granted without much willingness to ever get to the bottom of the arguments that have been offered for it, and the professional norms that define what counts as scientifically responsible philosophizing consistently reinforcing this behaviour.

If a work that is supposed to have established physical closure is cited though, it is often one of David Papineau's (2001, 2002, 2009). Near the end of his seminal "The Rise of Physicalism," perhaps the most oft-cited stock-taking of what is regarded by many the empirical evidence for physical closure, and an important exception to the tendency of taking physical closure for granted mentioned in the previous paragraph, he passes a judgment on the epistemic standing of the closure thesis (which he calls the "completeness of physics"):

[1] Early versions of this paper have been presented at the conference Mind and Metaphysics: a Conference in Honour of Howard Robinson at CEU Budapest, at the workshop Phenomenology of Agency and Free Will at the University of Fribourg. I have benefited greatly from the comments I got on both occasions. Comments are welcome at peter@szazholdas.hu.

> Of course, as with all empirical matters, there is nothing certain here. There is no knock-down argument for the completeness of physics. You could in principle accept the rest of modern physical theory, and yet continue to insist on special mental forces, which operate in as yet undetected ways in the interstices of intelligent brains. And indeed there do exist bitter-enders of just this kind […] However, I see no virtue in philosophers refusing to accept a premise which, by any normal inductive standards, has been fully established by over a century of empirical research. (Papineau 2001: 32-3.)

Not everybody shares this view, however. Lawrence BonJour, for one, writes:

> But why is the principle of causal closure itself supposed to be so obviously correct? Clearly this 'principle' is not and could not be an empirical result: no empirical investigation that is at all feasible (practically or morally) could ever establish that human bodies, the most likely locus of such external influence, are in fact never affected, even in small and subtle ways, by non-material causes. We are told that scientists accept this principle, and often that most philosophers accept it as well. But do they have any compelling reasons for such acceptance? Or is this vaunted principle nothing more than an unargued and undefended assumption—a kind of intellectual prejudice, in the literal meaning of the word? (BonJour 2010: 6.)

In this paper I will consider possible arguments for physical closure. In fact, I will attempt a rather comprehensive review of the types of arguments that could be tried to establish it. All of the particular arguments I will consider have actually been given by philosophers. So it would not be fair to say with BonJour that the closure principle is "nothing more than an unargued and undefended assumption." But I will argue that none of these arguments are any good. I will argue way more than that there is still room for doubting the closure thesis, in the manner of the "bitter-enders" Papineau despises. My aim is to show that so far we have come nowhere near to justifying it, by the normal standards of empirical science, and the prospects that we ever will are also dim. On the whole, the extensive reliance on this enormously significant thesis stands in stark contrast with the lack of sound argument or serious empirical evidence for it. So its epistemic standing is rather like BonJour sees it, after all. Hence the title of the paper.

2 The meaning of the closure thesis

Before we could start, I think I should justify the way I rendered the closure thesis. In the very first sentence of this paper, I formulated it as follows:

> Physical closure (PC): *Events or states of affairs that are not, or cannot be reduced to, physical events or states of affairs (if there are such events and states of affairs at all), make no difference to the course of physical events, or to the succession of the physical states of the universe.*

Admittedly, this is an unusual way of stating the thesis. This formulation is unusual mainly because it includes an explicit reference to whatever is outside the physical domain (tentatively, at least, since it is possible that there is nothing there really). Most formulations in the literature are restricted to be about the causal or nomological organization of, or the explanatory relations that hold in, the physical domain itself. See, for example, Papineau's formulation of the thesis from his "The Causal Closure of the Physical and Naturalism" in the *Oxford Handbook of Philosophy of Mind*:

> (P): *[E]very physical effect has a sufficient immediate physical cause in so far as it has a sufficient immediate cause at all.* (Papineau 2009: 59)

One can identify several features of this formulation which are designed to fend off complications which arise, I believe, because it aims to capture the idea of physical closure by attributing an intrinsic feature to the physical domain that supposedly amounts to it. A little consideration shows that it succeeds in some respects, but fails in others.

(P) is clearly designed to avoid the complication caused by the *transitivity of causation*. If we took out the qualification that we are to consider *immediate* causes only, (P) so altered would be compatible with the interactive dualist suggestion that some physical events are caused by non-physical mental events which themselves have sufficient physical causes (for a discussion of such cases see Lowe 2000).

It is also clear why (P) contains the clause that if a physical occurrence has a cause, it has a *sufficient* physical cause. For contrast, consider a similar formulation from Jaegwon Kim without this clause, (K): *If a physical event has a cause at t, it has a physical cause at t* (Kim 1993: 360). It would not be implausible from the interactionist's part to suggest that whenever a non-physical mental cause interferes in the physical causal order, it cooperates with a physical cause: in such a case neither the physical not the mental part of the complete cause is sufficient in itself, but each is necessary. In such a setup, interactionism is compatible with (K): even if a physical event has a necessary mental cause at t, it also has a physical cause at t without which the mental cause is not sufficient.

The conditional formulation of (P) – *in so far* (*i.e., if*) a physical event has a sufficient cause, (*then*) it is a physical sufficient cause – can accommodate the possibility that a Big Bang type cosmology may adequately capture the great outlines of the history of the universe, in which case uncaused physical initial conditions might have obtained at the beginning. For contrast, consider a formulation from David Chalmers, (C): *Microphysical causation and explanation seem to be autonomous, in that every physical event has a physical explanation; the laws of physics are sufficient to explain the events of physics on their own terms* (Chalmers 1996: 45). (C), being about the explanation of *every* physical event without qualification, seems to ignore this complication.

The same conditional formulation takes care of the possibility that the fundamental laws of physics, or the physical causal order, may turn out to be *indeterministic*: it takes care of the possibility that a particular physical event might not have a sufficient cause, which is the case if its physical causes fall short of necessitating that it will occur. (P) states that a necessary but insufficient physical cause cannot be completed to be sufficient jointly with the contribution of a non-physical mental cause, such as an immaterial mind controlling quantum indeterminacy, as it is suggested on the interactionist view which in this paper will be called *Eccles interactionism*.

Note, however, that an interactive dualist might suggest that non-physical mental causes operate in the room for manoeuvre created by quantum indeterminacy but stop short of fully controlling which of the physically possible outcomes should occur in such a situation. It is conceivable that non-physical mental causes only alter the probabilities of the possible outcomes of the collapse of the wave function (the reduction of the state-vector), assuming for the moment that a collapse-realist interpretation of quantum mechanics is true, but do not set the probability of any of the possible outcomes to 1. (P) clearly fails to rule out this possibility. Regardless of what will turn out to be the correct interpretation of quantum mechanics, a conceptually satisfactory formulation of the closure thesis should be incompatible with such a suggestion.

There is the further complication that some interactive dualists suggested that, rather than filling in the gaps in the physical causal order, the causal role of the non-physical mental is to make the difference between a measurement situation and a non-measurement situation, that is, to disrupt the smooth and deterministic evolution of a relevant quantum state governed by the dynamical law (the time-dependent Schrödinger equation) and induce a collapse. It would involve also "posing the probing question," that is, in the language of the formalism, to determine the base, according to which the pre-collapse mixed state will be expanded as a linear combination of eigenstates, and probabilities will be assigned to possible measurement outcomes that the eigenstates represent on the ground of their relative weights in the linear combination. In this paper, this view, tied to a specific interpretation of quantum mechanics, will be called *von Neumann interactionism.* (P) fails to forbid von Neumann interactionism. Again, even if the interpretation of quantum mechanics that involves this kind of interactionism is not very popular any longer, a conceptually adequate formulation of the closure thesis should not be compatible with it.

(P) has merits and shortcomings. I only used it to demonstrate that we have a pretty clear idea how to assess the merits and shortcomings of a formulation of the closure thesis. What we need to do is to consider conceivable interactionist scenarios and see whether it is strictly speaking incompatible with all of them. Any interactionist scenario we could think of would be relevant. Whereas, if a scenario does not amount to interactionism, as it is the case with the supposition of physically uncaused initial physical conditions under a Big Bang type cosmology, the formulation of the closure thesis should be compatible with it. It seems that what we are after is a formulation of the thesis that renders it equivalent to the negation of interactionism.

This needs some refinement, though. Some philosophers would classify the supposition that the physical effects of mental causes are overdetermined by the mental cause and an independent physical cause, each of which are sufficient to bring about the effect on their own, as form of interactionism, whereas this supposition is usually not considered to be in violation of physical closure, presumably because the redundant mental cause makes no difference to the physical causal order.

(PC) is essentially a negation of interactionism with this refinement.

Since almost every formulation of the closure thesis I have come across aims to state something positive about how the physical domain is intrinsically, rather than stating something negative about the relation between the physical domain and whatever is outside of it (i.e., that the latter does not influence the former), I suspect most philosophers think that it needs to be so for some reason. The only such reason I can think of is that they believe that there is something about how the physical domain is causally or nomologically or explanatorily organized in itself that entails how it is with respect to whatever is outside of it (i.e., that it is closed to any influence coming from it), and the proper formulation of the closure thesis should make this intrinsic feature of the physical domain explicit.

I don't see the need for such a restriction. Even if there was an intrinsic feature of the causal-nomological fabric of the physical domain which prohibited external influence, there would be nothing wrong about conceptualizing closure in terms of a lack of external influence. Of course such a formulation does not reveal what it is about the physical realm intrinsically that warrants against external influence. But perhaps there isn't such a thing. I can't think of any other candidate intrinsic feature of the physical domain that could entail closure than something about the laws of physics. But the physical realm can be closed even if closure is not entailed by anything about physical laws. It could also be known to us. Suppose, for example, that we have strong reason to believe in physicalism independently of physical closure. Then we would have strong reason to believe in physical closure even if the study of physical laws yielded no evidence for it. Closure would then arguably be a quite trivial thesis, uninteresting besides physicalism. But here is another, perhaps more interesting possibility. Suppose that we learn that a deterministic approach to quantum mechanics, the Bohmian interpretation, or one of the many versions of Everettianism, is true. Then, supposing that we have unlimited experimental capabilities to study the sequence of physical states that obtain at the relevant places (such as working human brains), proving closure would undoubtedly be an in principle viable empirical research project. We could do it by checking whether there is evidence for external influence in the actual course of physical events. Any physical event that deviates from what is predicted by the conjunction of earlier conditions and the laws of physics, which at the fundamental level would then be deterministic so the prediction would be unambiguous, would be a candidate for being such evidence. A persistent null-result could amount to a strong direct inductive justification for the closure thesis, even if the study of physical laws did not yield such an evidence. (Wouldn't the fact that physical laws hang together intrinsically by deterministic laws constitute evidence for closure in itself? No, because there is still the possibility that when the physical domain is considered not only in itself but also together with whatever is outside it, the deterministic laws that hold in the former case turn out to be ceteris paribus laws in the second.) In sum, it could conceivably be a brute empirical fact that external influence just never takes place, even though it could, for all the features of physical laws.

So I think using a negative, relational formulation like (PC), rather than a positive, intrinsic formulation like (P), is fully justified.

There is of course still the worry about the exact meaning of the adjective ‘physical’ in any formulation of the closure thesis. It has been argued that the uncertainty about the meaning of ‘physical’ is of the sort which would make the metaphysical thesis of physicalism either vacuous or certainly untrue (Crane and Mellor 1990). If it is so, it certainly has a bearing on trying to delineate an interpretation of the closure thesis that is both non-vacuous and justifiable. For the purposes of this paper, however, I will just assume, that the uncertainty about the meaning of ‘physical’ does not affect the justifiability of the closure thesis negatively. I propose to assume, for the sake of discussion, that the methodological and the ontological definitions of the physical (anything that is the subject matter of physics vs. anything that belongs to, or reducible to, a set of paradigm physical particulars and properties) coincide (or will coincide when physics is completed). Similarly, I propose to assume that no difficulty arises for the justifiability of the closure thesis from Hempel’s dilemma. In particular, I am happy to assume, for the sake of discussion, that all features of current physics that have been appealed to by the proponents of the closure thesis, assuming that these features justify their thesis, will be retained in all further editions of physics. All these assumptions, I believe, only benefit my opponent, and I am confident that there is no scientific case for physical closure even with these assumptions in place.

3 How to justify the closure thesis?

There are two classic arguments for physical closure, one due to Leibniz, the other due to Thomas Huxley, which, I believe, can be regarded as the two paradigm arguments that any possible empirical reasoning aiming to establish physical closure is bound to mimic in certain important respects.

Leibniz, of course, had reasons independent of physical closure to deny interaction between the mind and the body. Nevertheless, he seems to have interpreted Descartes’s interactionism as a failure of appreciating that conservation principles entail the closure of the physical realm. Leibniz’s argument makes use of a conservation principle for the discovery of which he claims credit, the conservation of (translational or linear) momentum (Woolhouse 1985). As Leibniz emphasizes, momentum has directionality, it is a vector we would now say. Descartes was aware of a closely related conservation principle, the conservation of the quantity of motion, mass times speed, which is a scalar, momentum deprived of its directionality. As it contains less information about the physical system in question than momentum, its conservation puts less constraint on its evolution, and this is what led Descartes astray, according to Leibniz, to thinking that there is room in physics for the non-physical mind to interfere (Leibniz 1705/1989). To be more precise, Leibniz thought the conservation of momentum ruled out interactionism *in conjunction* with the conservation of kinetic energy, whereas the conjunction of the conservation of kinetic energy and the quantity of motion did not have this consequence (Leibniz 1691/1896, 1695/1896).

Huxley, on the other hand, instead of studying what the laws of physics entail, studied the explanation of “the physical processes of life.” Life, including mental life, is the place where one would expect to find evidence against physical closure if there was any, in the form of physical events which resist explanation in the manner normal physical phenomena are explained, that is, in terms of previous physical events and states, and physical laws. We might interpret Huxley as advocating the closure thesis as an inductive generalization of the success of physical explanations for some physical processes of life. The thesis that all such processes “are capable of being explained in the same way as other physical phenomena,” leaving no room for extra-physical causes to interfere, has “grown in force and extent of application”, he claims. It is also reasonable to think that if it was false, empirical attempts to demonstrate its falsehood would be bound to succeed sooner or later, but, as Huxley insists, it “successfully repelled every assault which has been made upon it” (Huxley 1874).

I can think of no other way of justifying the thesis of physical closure than either by showing that external influence in the course of physical events is nomologically impossible, or by demonstrating through direct empirical research that external influence is in fact not happening.

As far as the first way is concerned, it could be more generally spelled out as an attempt to demonstrate that there is something about how physical events and physical states hang together intrinsically that entails how the physical domain is with respect to anything that is outside it, i.e. that it is closed with respect to any influence that would come from the outside. But physical events and states hang together through the laws of physics. So what this thesis about the intrinsic organization of the physical domain comes to is that the laws of physics entail closure. Leibniz thought, as many do today, that it is the conservation principles that entail closure, but it doesn't have to be so. It could be any law, or the conjunction of any laws. It could even be the general character of some laws (e.g. that they are deterministic). I label this way of justifying physical closure the 'D-way,' because it is a deduction of the closure thesis from particular physical laws or from some generic feature of some physical laws. There are three broad claims about physical laws that have been made by proponents of the closure thesis that need to be considered:

(D1) Conservation principles entail physical closure.

(D2) The dynamical laws of physics are deterministic, and this entails physical closure.

(D3) In case the dynamical laws of physics are probabilistic, this entails physical closure all the same.

There is a further claim which is not only about physical laws, but belongs with this family of deductive claims, and needs to be considered. It is concerned with closure with respect to external influence that would come from non-physical minds that operate through brains (or whole bodies) with which they are paired, by affecting the course of neural states that obtain in those brains (or whole bodies). This claim is essentially that even if the fundamental dynamical laws are indeterministic, the resulting indeterminacy does not propagate to the neural level; the laws of the dynamics of neural states is deterministic, and this forbids this kind of non-physical mental influence:

(D4) The dynamics of neural states is deterministic, and this entails closure with respect to non-physical mental influence.

The other way can be spelled out as studying the actual course of physical events, looking for irregularities in it, and if none is found, claiming that there is empirical evidence that external influence is not happening. This type of argument also has to do with physical laws because they dictate the regular course of physical events with which the actual course is compared to see if there are irregularities. But it doesn't say that there is something about physical laws, something about how physical events and states hang together intrinsically, that entails closure. This is rather a direct, empirical way of justifying physical closure. This way of justifying closure is bound to be inductive, since there is no way of checking every single physical event to verify that they all are regular. So I propose to call it the 'I-way.'

These two types of argument can be combined. They are in fact combined by Papineau who argues, as a first step, that the conservation of energy entails the impossibility of *spontaneous* external influence. Only law-governed external forces, whose force laws could be used to make sense of a non-physical potential energy, are allowed by the principle of the conservation of energy, he claims, because then decreases or increases of non-physical potential energy will compensate for the energy given or taken away from the physical system when the non-physical external forces bring about a physical change. As a second step, he claims that a century of empirical physiological research failed to find such law-governed external forces in living organisms, which is the place to look for such forces, so, by normal inductive standards, physical closure is vindicated (Papineau 2001).

Although, as it will turn out shortly, this particular argument is not sound, I will argue that physical closure could in principle be justified only by combining the two broad types of argument, the D-type and the I-type. It is because there are two types of conceivable external influence, and only a D-type argument could rule out one of them, and only an I-type argument could be used to rule out the other. The relevant distinction, however, is not between spontaneous and law-governed external influences, but the one I am about to introduce.

4 Two types of external influence and the plan for the rest of the paper

I propose to consider two sorts of conceivable external influence in the course of physical events, or in the succession of physical states.

It is conceivable that external factors sometimes influence the course of physical events in such a way that the deviant course of physical events is just as consistent with the laws of physics and the physical conditions that obtained before the external influence took place as the normal course would have been, that is, the course that would have obtained in the lack of external influence. This of course requires physical determinism to be false, but, for all we presently know, it may be false. I propose to call this kind of external influence *law-conforming.*

It is also conceivable, however, that the external influence is not law-conforming. I propose to call such an influence *overriding.*

To some extent it is a misnomer. If such an influence exists, we are not bound to regard the situation as one in which a law is being counteracted. We can regard it as one which reveals that a law which holds perfectly well for physical systems properly isolated from external influence is really a ceteris paribus law when applied to the same systems placed in an environment in which external influence may take place. If the relevant physical laws are regarded as ceteris paribus laws, they are not strictly speaking overridden by what I have just proposed to call an overriding external influence. Nevertheless, I will stick to this terminology. Of course, in the definition of law-conforming influence the relevant laws are taken at face value, that is, without ceteris paribus clauses.

With this distinction at hand, I am ready to lay out my plan for the rest of this paper.

In the next section I will argue for the relatively obvious point that D-type arguments cannot rule out overriding external influence, whereas I-type arguments cannot work against law-conforming external influence. As a justification for the closure thesis would need to rule out both law-conforming and overriding external influence, it should then consist of both a D-type argument against law-conforming external influence and an I-type argument against overriding external influence.

In the sections after the next I will argue against the D-type claims I am aware of, D1, D2, D3 and D4, as applied to law-conforming external influence. So the target claims I will try to refute will be:

(D1) Conservation principles are inconsistent with law-conforming external influence.

(D2) The dynamical laws of physics are deterministic, and this is inconsistent with law-conforming external influence.

(D3) In case the dynamical laws of physics are probabilistic, this is just as inconsistent with law-conforming external influence as if they were deterministic.

(D4) The dynamics of neural states is deterministic, and this is inconsistent with law-conforming non-physical mental influence.

Finally, I will examine what it would require to make sure the I-way, that is, by direct empirical methods, that there is no overriding external influence in the course of physical events. I will argue that we have not done what it would take, and there is no guarantee that we will ever be able to do what it would take.

5 No I-type argument (inductive direct empirical argument) against law-conforming external influence, no D-type argument (deductive argument form the laws of physics) against overriding external influence

The first conjunct in the title of this section is quite obvious. Law-conforming external influence is by definition undetectable by an I-type methodology. An I-type argument would be an inductive argument on the ground of the null-result of a methodologically adequate empirical search for nomological irregularities in the course of physical events (at places where it is relevant to look for such irregularities, e.g. in working human brains). This method cannot work to rule out law-conforming influence, since the events that would be the effect of such an influence would not be nomologically irregular, so they would be invisible for the detection method.

The second part of the claim is quite obvious too, but seeing it requires a somewhat more substantive consideration. The existence of overriding influence in the course of physical events would entail that the relevant laws which appear to be overridden are ceteris paribus laws. The question to consider is whether it is compatible with what we know about the laws of physics to suppose that they are possibly ceteris paribus laws. (If not, then this would be a D-type argument against overriding external influence.) I can offer two simple arguments for the reasonableness of regarding the laws of physics as ceteris paribus laws, one theological, and one drawing on a simplified model of the experimental methodology of physics.

Here is the theological argument.

Imagine that there is a God who could interfere in the course of physical events, but for some reason He freely refrains from doing so (or at least refrains every time when there is the slightest chance that His intervention would be detected by the creatures in the world who are endowed with free will). (To be vivid, we may suppose that God's reason to choose to restrain Himself in such a way has to do with His ideas about the sort of world in which free will can be exercised rationally and responsibly, i.e., being potentially fully aware of what the consequences of one's choices will be, and being certain that the consequences will obtain, but this is not essential to the story. Something very much like this has actually been suggested in the context of the argument from evil by Bruce Reichenbach (1976).) There seems to be no a priori ground to rule out such a suggestion. Little consideration shows that such a suggestion cannot be ruled out by anything about the laws of physics either.

There is nothing incoherent in the thought of two possible worlds which are identical except that there is a freely self-restraining God in one of them, but not in the other. Since physicists come to discover the laws of physics by studying the course of physical events and the succession of physical states, ideally, the physicists of these two worlds would come to exactly the same physical laws (maybe modulo the choice of some different but isomorphic mathematical structures to model physical causal-nomological structures). Unknown to the physicists, in one of these worlds the true laws of physics of course bear the tacit clause "unless God chooses to interfere." But even in the Godless counterpart world, there can be nothing about the laws that is inconsistent with such a suggestion, given that they have exactly the same empirical grounding as in the God-containing world.

Now apply this to our world. Either there is a God in our world who could override the physical laws, or there is a possible world physically indistinguishable from ours in which God could override the physical laws. Consequently, there can't be anything about the physical laws of our world that would entail that its physical domain is closed with respect to overriding external influence.

And now the argument from the experimental methodology of physics. This argument has been given, in much more detail, by Robert Bishop (2006). So a brief statement of the core point, drawing on a simple model of how we read off physical laws from experiments, will do for now.

It is an essential part of the methodology of physics that in the experiments from which we read off the laws of physics we systematically try to screen off the noise created by causes which are not the one whose effects are being studied at the time. Consider the following example.

When Galileo established that different objects dropped from the same height take the same time to hit the ground, he was interested in the effect of Earth's gravitational pull, for which reason he had to use relatively dense and compact objects to screen off the effect of drag. Doing the experiment with feathers would not have yielded the result he was interested in. The ideal way to perform the experiment would have been to do it in vacuum, to shield drag off completely. So the physical law *"objects of different weights and material compositions dropped from the same height take the same time to hit the ground"* is really a ceteris paribus law, the ceteris paribus condition being *"unless drag affects their movement differently."*

Once we have bits of Newtonian dynamics in place, we can easily do away with this ceteris paribus clause. We eliminate it by supplementing a force-law for drag, and plugging it into Newton's second law alongside the gravitational force-law. The amended law, stating that dropped objects will move according to Newton's second law, in conjunction with the force-laws for gravitational pull and drag, will require no ceteris paribus clause about drag to be added. This illustrates a general point. As long as the ceteris paribus clauses attached to physical laws refer to the possible effects of further physical causes, in principle, we can get rid of them by adding further bits of physical theory, which will result in more accurate laws that don't need to be qualified by the such clauses.

Now consider the possibility that someone, let's call him Giovanni, takes it into his mind that he wants to mess up Galileo's experiments. He hangs out on Campo dei Miracoli, sneaks up to the Leaning Tower when he sees Galileo scaling the stairs, and when Galileo drops his objects, he tries to catch them mid-air. Strictly speaking, the physical law *"objects of different weights and material compositions dropped from the same height take the same time to hit the ground"* can be read off from the experiments only qualified by the ceteris paribus condition *"unless Giovanni catches them mid-air."* This ceteris paribus clause can be discharged the same way as the clause referring to the possible effect of drag was in the previous case only if there are bits of physical theory to be supplemented which fully cover the processes in Giovanni's brain that make, or don't make, him stick out his hand and catch the falling objects in a given experiment.

If either reductive physicalism or epiphenomenalism is true, or if it is true that non-physical mental causes always operate redundantly alongside a physical cause which is sufficient in itself to bring about the effect, then the required bits of physical theory, at least in principle, can certainly be supplied to cover this case, too. But if reductive physicalism or epiphenomenalism or overdetermination is true, physical closure is trivially true. The interesting question is whether we can justify physical closure without assuming any of these three. As long as we read-off the laws of physics from experiments in controlled environments to fend off unwanted interference by causes which are not the one whose effects are being studied, it is reasonable to assume that the laws we obtain this way are ceteris paribus laws, and we cannot just assume that the ceteris paribus clauses are always eliminable without remainder by supplementing other bits of physical theory, since that would amount to assuming that either reductive physicalism, or epiphenomenalism, or systemic overdetermination is true.

What these arguments establish is that there is nothing we know about physical laws that would rule out the possibility that they are ceteris paribus laws, their ceteris paribus clauses making reference to possible interventions from outside the physical realm. This is to say, the laws of physics do not entail that there is no overriding external influence in course of physical events, or, in our jargon, there is no D-type argument against overriding external influence.

Previously we made the observation that there is no I-type argument against law-conforming external influence. A justification for physical closure would require warrant against both law-conforming and overriding external influence, so *a justification for physical closure would have to consist of a D-type argument against law-conforming external influence and an I-type argument against overriding external influence*. If either of the two is missing, the closure thesis is unjustified. In the rest of the paper I will argue that in fact both are missing.

6 Do conservation principles entail closure with respect to law-conforming external influence?

I will examine three specific claims about conservation principles and physical closure. In line with the discussion of the previous section, throughout this section we will be concerned with closure with respect to law-conforming external influence, even if it is not explicitly mentioned.

The claims to be considered are these:

(D1A): *The conservation of energy entails physical closure.*
(D1B): *The conservation of energy entails closure with respect to spontaneous external influence.*
(D1C): *The conjunction of conservation principles entails physical closure.*

My aim in this section is to show that none of these claims have any real support in physics.

A) Does the conservation of energy alone entail closure?

Daniel Dennett, for one, seems to believe this is the case:

> [T]he principle of the conservation of energy is apparently violated by dualism. This confrontation between quite standard physics and dualism has been endlessly discussed since Descartes' own day, and is widely regarded as the inescapable and fatal flaw of dualism (Dennett 1991: 35).

Paul Churchland recently wrote that both the conservation of energy and the conservation of momentum *alone* entail closure:

> [A]ny position that includes non-physical elements in the causal dynamics of the brain must violate both the law that energy is neither created nor destroyed, and the law that the total momentum in any closed system is always conserved (Churchland 2014: 56).

He goes on explaining why he believes this is so:

> In short, you simply can't get a change in any aspect of the physical brain (for that would causally require both energy changes and momentum changes) save by a compensatory change in some other physical aspect of the brain, which will thereby lay claim to being the cause at issue. (Ibid.)

We will have to start our discussion of whether the conservation of energy entails closure by constructing an argument to this effect on behalf of the proponent of (D1A). I suggest we use this very condensed last sentence cited from Churchland as a starting point.

Churchland is explicitly committed to the view that "a change in any aspect of the physical brain […] would […] require […] energy changes." This will serve as our first premise.

The second half of the same sentence reveals that he is also committed to the view that such changes would require "a compensatory change in some other physical aspect of the brain." I suppose that that change, in order to be compensatory, would need to be an energy change, too. Requiring that it should be "in some other physical aspect of the brain" suggests that Churchland thinks that the brain is a conservative (energy-conserving) system. Since the brain is in rich physical interaction with its environment, no doubt involving transfers of energy, this suggestion is immensely implausible. (For one thing, the brain gives off so much heat to its environment that it famously led Aristotle to think it was its main function. Giving off heat is energy transfer.) But we can replace this implausible premise with the much more plausible one that an energy change in some part of the brain would require a compensatory change in some other (presumably nearby) part of the physical world, which is to say that the physical world as a whole is a conservative system.

These two premises do not yet get us a valid argument for closure:

(1) For an external influence to take effect on the brain, it would be required to bring about an energy change in the brain.
(2) The principle of the conservation of energy is true, and it means that the physical world is a conservative system.

These two premises fall short of implying closure because the conjunction of them can be satisfied if there is an energy change in some other part the physical world that compensates for the energy change in the brain brought about by an external influence affecting the brain. Churchland, however, adds a third premise in the last bit of the sentence we are considering along the following lines:

(3) If there is such a compensatory energy change in some other (nearby) part of the physical world, we are bound to identify it as the cause of the change in the physical state of the brain we are considering, thereby dropping the hypothesis that it was brought about by an external influence.

Premises (1-3) together entail closure. If they are true, together they amount to what one might call an indirect proof that there is no external influence in the succession of the physical states of the brain. One starts off by hypothesizing that such an influence causes a change in the physical state of a brain, and then consideration of the three premises forces one to drop this hypothesis. However, all three premises are doubtful, especially (1) and (3). The most obvious point of attack is (3). There are countless physical examples when a cause brings about an energy change in some physical subsystem and a compensatory energy change of the opposite sign in another physical subsystem. When a driver steps on the brake pedal in a moving car, for example, this brings about a decrease in the kinetic energy of the car, which then is compensated by the breaks getting hot and giving off heat to their environment, which is an increase of the mean kinetic energy of the atoms and molecules in and around the breaks. It would be totally unreasonable to assume that the increased temperature of the brakes and the air in their vicinity caused the car to slow down, rather than the driver's stepping on the brake pedal. Of course, stepping on the brake pedal and the workings of the brake mechanism involve energy changes in certain body parts of the driver and in some car parts. But this is not the energy change that compensates for the change of the car's kinetic energy. If a physical cause can bring about a redistribution of energy in a physical system, I see no principled reason to suppose that a non-physical cause couldn't.

Dropping (3), we need to alter the remaining premises to get an argument for closure. Here is one possibility:

(1') For an external influence to bring about a physical effect, it would be required to change the amount of energy in the physical world.

(2) The principle of the conservation of energy is true, and it means that the physical world is a conservative system.

Again, these premises entail closure, but (1') is highly suspicious, and there is room for doubt about (2) as well, even if we disregard, for the sake of discussion, the complications that arise in the general relativity. It is not entirely clear that the principle of the conservation of energy means that the physical world as a whole is a conservative system. Experimental support for the principle comes from the examination of physical systems properly isolated from their environments. Theoretical support for the principle comes from the fact that it is entailed by the highly plausible symmetry requirement (the homogeneity of time), taken together with a highly abstract but empirically well supported variation principle (Hamilton' principle, see below towards the end of Subsection C). But it is unclear what the principle would require in a larger world, in which the physical world is just a subsystem. The principle is empirically justified for causally isolated systems and the proponent of the closure thesis cannot just assume that the physical world is causally isolated from the rest of the world without begging the question against the interactionist. It seems perfectly conceivable that there are also law-governed non-physical forces with physical effects in such a world. As Papineau considers in his argument for (D1B), to be looked into in a minute, on the ground of such force-laws, potential energy could be attributed to the non-physical subsystem. The equations of completed physics in this world would presumably be required to include factors corresponding to these non-physical yet law-governed forces, and the symmetry consideration (the homogeneity of time) that entails the conservation of energy would be applicable to these equations. Plausibly, when imagining such a world, we could think of the principle of the conservation of energy as requiring that the whole world is conservative system, even though the current empirical evidence we have for the principle is based entirely on the examination of isolated physical systems. (For differing views on this issue see Montero 2006 and Koksvik 2007.)

But the truth of (2) is our lesser concern, because (1') is certainly false. As we have seen when we discussed premise (3) of the original Churchland-style argument, nothing prevents us from considering it possible that when an external influence has a physical effect, it redistributes energy in the physical world.

At this point, however, one could suggest that it still might be the case that in order to bring about a change in the physical world that involves the redistribution of energy in it, one would need to supply or take away energy at some point. In the above example with the braking car, the process redistributes energy; basically, it turns the kinetic energy of the car into heat. But the event that triggers it, the driver stepping on the brake pedal, involves energy changes in the driver's body and in the brake mechanism. Maybe this observation can be generalized along the following lines: (a) in order to trigger a redistribution of energy in the physical world one has to bring about a change in a particular physical system, which will serve as the event that triggers the chain of events that results in the redistribution of energy in the physical world, and (b) to bring about this change in a particular physical system, the amount of energy in it must be changed. The latter part, (b), is the supposition underlying premise (1) in the original Churchland-style argument, which was that in order to achieve a physical change in the brain, the non-physical mind would have to alter the amount of energy in it.

But even this supposition lacks any support from physics. Already in 1925, C.D. Broad brought to the attention of the philosophical community the simple physical fact that a physical cause is not required to bring about an energy change in a physical system in order to affect it. So it is unreasonable to assume that a non-physical cause would be required to do so.

Broad demonstrated it on the example of a pendulum. I am not hoping to achieve an improvement on his presentation overall, but I can offer an even simpler example: a space station orbiting a planet on a perfectly circular orbit. It is simpler because in the case of the pendulum there are two forces to consider, the pull of gravity, and the pull of the rope. In the case of the space station orbiting a planet there is only gravity.

Clearly, the planet's gravitational pull affects the space station. This is what keeps it on orbit. If it was miraculously turned off, the space station would fly off along a straight line, tangential to its original circular orbit. The gravitational pull of the planet accelerates the space station, the acceleration vector always pointing towards the planet. But no energy is thereby communicated to the space station. The energy communicated to it would be equal to the work performed on it by the gravitational force. But the work performed by a force on a body is defined as the scalar product of the force vector and the vector representing the displacement of the body. If the space station does perfect circles around the planet, the displacement vector is perpendicular to the force vector at all times, which means that their scalar product is always zero. So no work is being done, no energy is being communicated.

Getting back to the example of the moving car, we can exploit an analogous setup to construct a scenario in which the initial event of the causal sequence that leads to the engaging of the brakes and eventually to the stopping of the car involves no energy change. Suppose that in a roof box mounted on the car, a charged particle is emitted from a source in the direction of the car's movement, say, every second. There is also a detector, placed somewhat aside from the charged particle's path. Suppose further that at some point the car reaches an area where there is an external static magnetic field. As the car gets there, the magnetic field affects the movement of the charged particle emitted in that second. The particle diverts just so that it hits the detector. The detector is hooked up to the brake mechanism: when it goes off, the brakes engage. In this scenario, the physical event that triggers the braking, and the turning of the car's kinetic energy into heat, is the emitted particle getting diverted by the magnetic field. This is analogous to the space station case because the force that diverts the particle from its original path to hit the detector is determined by the Lorentz force law. The Lorentz force law says that the force vector that affects a charged particle moving in a magnetic field is charge times the vector product of the particle's velocity vector and the magnetic induction vector. The vector product of two vectors is always perpendicular to both, so the force vector will be perpendicular to the direction of the particle's movement throughout the whole movement, which entails that the work that the magnetic field performs on the particle is zero, which is to say that no energy is given to, or taken away from, it.

So not even the underlying general principle behind premise (1) that in order to affect a physical system the amount of energy in it needs to be changed is supported by physics. It doesn't have to be so in the case of physical causes, so there is no reason to assume that it has to be so if the cause is non-physical.

Time to sum up.

We have considered an argument on behalf of a theorist who believes that the only physical law that needs to be invoked as a premise to a deductive argument for physical closure, which was based on a passage cited from Churchland. The choice of the passage, and so the choice of the argument to discuss, was arbitrary. But there is a lesson from the discussion that generalizes to other possible arguments from the conservation of energy to closure. The lesson is that neither the stronger premise that in order to bring about a physical effect, an external cause (such as a non-physical mind) would be required to change the amount of energy in the physical world, nor the weaker premise that it would be required at least to change the amount of energy in the physical subsystem it directly affects (in the case of a mind, presumably the brain, even if only to trigger a redistribution of energy in the physical world), has any support from physics. Since I can't think of any argument on behalf of the proponent of (D1A) which does not involve one of these premises, either (1) or (1'), I am quite confident that physical closure cannot be established relying on the conservation of energy alone.

B) Does the conservation of energy entail closure with respect to spontaneous external influence?

This weaker claim has been made by Papineau in "The Rise of Physicalism." This is how he argues for it:

> The conservation of energy bears differentially on these two kinds of special forces [i.e. spontaneous vs. governed by a force-law]. It does seem inconsistent with the first kind of special force, a spontaneous special force. But it does not directly rule out the second, deterministic kind.
>
> Why should the conservation of energy rule out even a spontaneous special force? (Think of a spontaneous mental force that accelerates molecules in the pineal gland, say.) Why shouldn't such a force simply respect the conservation of energy, by not causing accelerations which will violate it? But this doesn't really make sense. The content of the principle of the conservation of energy is that losses of kinetic energy are compensated by build-ups of potential energy, and vice versa. But we couldn't really speak of a "build-up" or "loss" in the potential energy associated with a force, if there were no force law governing the deployment of that force. So the very idea of potential energy commits us to a law which governs how the relevant force will cause accelerations in the future.
>
> However, nothing in this argument rules out the possibility of vital, mental, or other special forces which are governed by deterministic force laws. After all, the conservation of energy in itself does not tell which basic forces operate in the physical universe. Are gravity and impact the only basic forces? What about electromagnetism? Nuclear forces? And so on. Clearly the conservation of energy as such leaves it open exactly which basic forces exist. It only requires that, whatever they are, they operate conservatively.

After the discussion of the previous subsection, it is immediately clear what is wrong with this argument. It involves the premise that a mental force could bring about an "acceleration" – I take that "acceleration" here stands for any change in the physical world, presumably in some part of the brain – only thereby bringing about a change in the energy level of the physical world, or at least of some part of the brain, what would need to be compensated by changes in non-physical potential energy, of which we can make sense only if the non-physical force in action is governed by a force-law. What it comes to is either (1) or (1') discussed in the previous subsection and found unsupported by physics.

C) Does the conjunction of all the relevant conservation principles entail closure?

Maybe conservation principles do a better job establishing closure if they are considered not one by one but taken together. This is what Leibniz believed to be the case. In particular, he believed that the conservation of kinetic energy and the conservation of momentum entailed closure. We will assess the general claim that all the relevant conservation principles taken together entail closure.

Initially, we will do it on the example of a ball-world of the sort that might have been a familiar image of physical reality in Leibniz's time. It consists entirely of balls and otherwise empty space. The only interaction between the balls are collisions, apart from collisions they traverse undisturbed in empty space. Both the balls and the collisions are metaphysical primitives. The collisions are assumed to preserve kinetic energy; in contemporary parlance we would say that they are perfectly elastic. The ball-world image of physical reality is of course far behind us. Nevertheless, it proves a useful instrument to demonstrate an essential point about conservation principles.

As a first step, to secure maximum degree of simplicity, let us consider a one-dimensional ball-world. It would consist of balls that are allowed to move along a straight line only. Let us suppose also that the balls do not rotate. It means that the conservation of kinetic energy and the conservation of linear momentum are the only two conservation principles to consider. The history of this world is made up by events such as balls moving uniformly along a straight line, and balls colliding with each other. Changes in any ball's movement occur only in the collisions. Again, for the sake of simplicity, let us suppose that there are only two-ball collisions, if two balls touch, neither of them touches a third ball. Let us zoom in on two balls colliding with each other.

Since it is assumed that the only interaction in this world is collision, and there are only two-ball collisions, a couple of balls colliding is a closed subsystem of this world, so the conservation laws are applicable. In this one-dimensional case, the two relevant conservation principles, the conservation of linear momentum and the conservation of kinetic energy determine how the two balls should emerge from the collision, given the way they moved beforehand. (If the balls move in empty space, and the collisions are perfectly elastic, kinetic energy is conserved throughout the entire history of this world.) The two conservation principles yield two equations in which the balls' post-collision velocities figure as variables. With the usual notations (indices 1 and 2 refer to the two balls, unprimed and primed variables represent pre-collision and post-collision states, respectively):

$$m_1 v_1 + m_2 v_2 = m_1 v_1' + m_2 v_2'$$

$$\frac{1}{2} m_1 v_1^2 + \frac{1}{2} m_2 v_2^2 = \frac{1}{2} m_1 v_1'^2 + \frac{1}{2} m_2 v_2'^2$$

Supposing we know the two balls' masses and their pre-collision velocities, this is a system of two logically independent equations for two unknown variables, v_1' and v_2'. It has a unique solution. There is only one way for these balls to emerge from the collision consistently with their pre-collision state and the conservation laws.

Do these two conservation principles entail closure with respect to law-conforming external influence in this one-dimensional ball world? Yes, they do.

Since the only interaction in which a ball can be involved is collision with another ball, between any two collisions, every ball is causally isolated, as far as physical causes are concerned. This makes the conservation principles applicable to single balls in flight. A non-physical mind couldn't tamper with them without braking both of the relevant conservation principles. Note that, in reality, a ball would be affected by at least the gravitational pull of every other ball, so it could not be regarded to be causally isolated between collisions, and therefore it wouldn't be appropriate to apply the conservation principles to the 'system' of a single ball in-flight, except maybe approximatively. But it shouldn't concern us as long as we consider this hypothetical world.

When a collision takes place, the situation is much the same. The conjunction of the two conservation principles entails the balls' post-collision movement, given their pre-collision movement, so if a non-physical mind would want to make them come out of the collision in any other way, it would bring about a situation which is inconsistent with the conjunction of the two conservation principles and the pre-collision movement of the balls.

So far so good from the perspective of the proponent of the closure thesis, but this one-dimensional ball-world is about the most complex system whose closure is guaranteed by the conjunction of these two conservation principles, as we shall see in a minute.

Consider now a ball-world exactly like the previous one, except two-dimensional. This world would be much like a set of billiard balls moving (gliding rather than rolling, since they are not allowed to spin) on a perfectly frictionless table in vacuum. Let us zoom in on a two-ball collision like we did in the previous case.

In two dimensions, the conservation of momentum yields two equations instead of one. The conservation of kinetic energy yields one equation as before. With the same notations (except that new indices x and y are introduced to represent two orthogonal directions in a two-dimensional Cartesian coordinate system, and, for the sake of simplicity, the masses of the two balls are assumed to be equal):

$$mv_{1x} + mv_{2x} = mv'_{1x} + mv'_{2x}$$

$$mv_{1y} + mv_{2y} = mv'_{1y} + mv'_{2y}$$

$$\frac{1}{2}m\left(v_{1x}^2 + v_{1y}^2\right) + \frac{1}{2}m\left(v_{2x}^2 + v_{2y}^2\right) = \frac{1}{2}m\left(v_{1x}'^2 + v_{1y}'^2\right) + \frac{1}{2}m\left(v_{2x}'^2 + v_{2y}'^2\right)$$

Since now both balls can move in two dimensions, we have four unknown variables: v'_{1x}, v'_{1y}, v'_{2x}, and v'_{2y}. This is one too many for three independent equations. There is no unique prediction on the basis of these two conservation principles for the post-collision movement of the two balls.

Now let us ask the same question as before: Do these two conservation principles entail closure with respect to law-conforming external influence in this two-dimensional ball-world? The answer is no.

The consideration for the situation between any two collisions is the same as it was in the one-dimensional case. But when it comes to how the balls emerge from a collision, a non-physical mind would have a marge of manoeuvre to influence the course of the ball-world without getting into conflict with the conjunction of these two conservation principles, since there is more than just one way to satisfy them, given the pre-collision states.

In fact, using a frame of reference that is in rest relative to ball 2 before the collision, the conservation of momentum entails that the vector-sum of the post-collision velocities of the two balls equals the pre-collision velocity vector of ball 1, if the two balls have equal masses (since in this frame the pre-collision velocity of ball 2 is a zero vector). Represented geometrically, it means that the three vectors, the pre-collision velocity vector of ball 1, and the post-collision velocity vectors of ball 1 and ball 2, form a triangle. The conservation of kinetic energy is an equation linking the squares of the lengths of these vectors. Through the Pythagorean theorem, it entails that the triangle formed by the three velocity vectors is a right triangle. But, as far as these two conservation principles go, the right angle of this triangle could be anywhere on a Thales' circle whose diameter is the pre-collision velocity vector of ball 1.

However, the conservation of angular momentum comes to the rescue of the theorist who wants to argue for closure from the conjunction of the conservation principles in this world. The conservation of angular momentum imposes a further constraint on how the balls should emerge from the collision.

Recall that we are considering a world in which the balls don't spin. If this is so, then the total angular momentum of the two-ball system we are considering is the sum of the two balls' angular momentum about the origin. Let $\boldsymbol{r}_1$and $\boldsymbol{r}_2$ denote the location of the centres of ball 1 and ball 2, respectively, whereas $\boldsymbol{v}_1$ and $\boldsymbol{v}_2$ are their velocities. (Bold letters denote vectors.) With these notations, relative to the origin of our frame of reference, the total angular momentum of this couple of balls is

$$\boldsymbol{H} = \boldsymbol{r}_1 \times m\boldsymbol{v}_1 + \boldsymbol{r}_2 \times m\boldsymbol{v}_2$$

(The operator × means vector product.)

Now the conservation of angular momentum means that the derivative of this formula is zero throughout the collision process:

$$\dot{\boldsymbol{H}} = \dot{\boldsymbol{r}}_1 \times m\boldsymbol{v}_1 + \dot{\boldsymbol{r}}_2 \times m\boldsymbol{v}_2 + \boldsymbol{r}_1 \times m\dot{\boldsymbol{v}}_1 + \boldsymbol{r}_2 \times m\dot{\boldsymbol{v}}_2 = 0$$

Since $\dot{\boldsymbol{r}}_1 = \boldsymbol{v}_1$ and $\dot{\boldsymbol{r}}_2 = \boldsymbol{v}_2$ (velocity is the derivative of location),

$$\dot{\boldsymbol{r}}_1 \times m\boldsymbol{v}_1 = \dot{\boldsymbol{r}}_2 \times m\boldsymbol{v}_2 = 0$$

given that the vector product of two parallel vectors is zero. From this, and the previous equation about the derivative of the total angular momentum, it follows that

$$\boldsymbol{r}_1 \times m\dot{\boldsymbol{v}}_1 + \boldsymbol{r}_2 \times m\dot{\boldsymbol{v}}_2 = 0$$

Given the conservation of linear momentum, any change in the linear momentum of ball 1 is always accompanied by a change in the linear momentum of ball 2 equal in size and opposite in direction, therefore,

$$m\dot{\boldsymbol{v}}_2 = -\, m\dot{\boldsymbol{v}}_1$$

at all times. With this substitution:

$$(\boldsymbol{r}_1 - \boldsymbol{r}_2) \times m\dot{\boldsymbol{v}}_1 = 0$$

Since neither $\boldsymbol{r}_1 - \boldsymbol{r}_2$, nor $m\dot{\boldsymbol{v}}_1$ is zero, this vector product vanishes only if $m\dot{\boldsymbol{v}}_1$ is aligned with $\boldsymbol{r}_1 - \boldsymbol{r}_2$, which means that the acceleration vectors of the balls are aligned with the line connecting the centres of the two balls.

Then, using a Cartesian reference frame whose x axis is the line connecting the two centres, the collision process decomposes into a head-on one-dimensional collision along the x axis, and an undisturbed movement of both balls along the y axis. (Neither of the two balls undergo any acceleration in the y direction.) So the problem is reduced to that of a one-dimensional collision, which, as we have seen, is predictable on the ground of the conservation of kinetic energy and the conservation of angular momentum. In other words, if we take into account the conservation of angular momentum, this two-dimensional ball-world is predictable. Leibniz would have been perfectly right, had he considered all three conservation principles instead of just two. If the frame of reference is at rest relative to ball 2 before the collision, the post collision velocity vector of ball 2 will be aligned with its acceleration vector, given that it has no initial velocity. It means that, in our geometrical representation, the line connecting the centres of the two balls when they collide marks out the place of the right angle of the right triangle formed by the velocity vectors on the perimeter of the Thales' circle. There is room for Cartesian dualist interference consistently with the conjunction of the conservation of kinetic energy and the conservation linear momentum, but not with the conjunction of all three conservation principles. Note, however, that this is dependent on the hypothesis that the balls don't spin. (To Leibniz's credit, that the forces that affect the balls attack along the line connecting their centres is a plausible hypothesis to make even if one is not aware of the conservation of angular momentum.)

Now we are in the position to observe a pattern: in order to exclude external influence that is consistent with the conjunction of the relevant conservation principles, the constraints they impose on the evolution of the system we are considering must outnumber the system's degrees of freedom.

A couple of non-spinning balls in two dimensions have four degrees of freedom, that is why the two conservation principles we considered first were insufficient to fix how they emerge from collision given their pre-collision state, and the conservation of angular momentum had to be called in. Appealing to further conservation principles, however, is not a strategy that can be continued indefinitely, for the simple reason that we run out of conservation principles very soon as the system gets more complex. When we consider systems even slightly more complex than the one we have just did, the number of the system's degrees of freedom easily outrun the constraints that can be imposed on the ground of the relevant conservation principles.

Consider, for example, the very same ball-world, except with balls that can spin. This means three extra degrees of freedom per ball. Since the spin of the balls contribute to the overall angular momentum, we cannot deduce from the conservation of angular momentum that the forces that accelerate the two balls attack along the line connecting their centres, and then we are back with the situation that the right angle of the triangle formed by the velocity vectors can be anywhere on the Thales' circle, as far as all three conservation principles go. Allowing three- or n-ball collisions, instead of just two at a time, would have a similar effect: too many degrees of freedom for the system to be constrained enough by the conservation principles to exclude external influence.

This is the general point I wanted to illustrate with these old-fashioned ball-worlds. Conservation principles are only concerned with aggregates of certain physical properties at different times in the evolution of a system, in consequence, they lack the required resolution, so to speak, to predict the details of how the system will evolve, provided that the system has more degrees of freedom than the number of independent equations that the conservation principles yield. The relevant system we are considering when we ask the question whether a non-physical mind could bring about a change in the course of physical events is the brain, taken together with everything with which it physically interacts. It is complex enough.

Haven't I cheated? Aren't all the ball-worlds we considered deterministic, after all? How can I claim that a non-physical mind could interfere in its course without breaking the laws of physics?

Well, on the classical dynamical description, yes, these ball-worlds, including the one in which the balls are allowed to spin, are all deterministic. Supposing that the classical dynamical description is correct, of course I cannot claim that a non-physical mind could tamper with these worlds without breaking any law. What I am claiming is that a non-physical mind could interfere with a complex enough world (like the two-dimensional ball-world in which the balls are allowed to spin) *without breaking the conservation laws*. The ball-worlds are deterministic alright on the classical description, and determinism of course implies that a Cartesian mind cannot make a difference to the course of a world's evolution without braking any physical law, which is to say that determinism implies closure with respect to law-conforming external influence. But the point is that, if a ball-world is complex enough, its determinism is *not implied by the conservation principles.* What the proponent of the closure thesis needs to establish closure with respect to law-conforming external influence is the determinism of the full dynamical description.

To reinforce this point, let us finally consider an altogether different, and somewhat more complex world. This world consists of n "balls," this time conceived of as massive point-like particles, interacting with each other throughout the entire evolution of the world. Each particle affects the movement of every other at all times, and the effect of the interaction only depends on how the particles are located in space. (Sets of particles interacting gravitationally or electrostatically would be examples for such a world. The two-particle case in which the forces that arise from the interaction attack along the line connecting the two particles and vary with the inverse square of the distance between them is of course the familiar Kepler problem.) We will briefly consider a classical, non-relativistic description of such an n-particle world.

The first thing to note about such a world is that our previous consideration about in-flight balls, which was valid in worlds in which balls interact with each other exclusively in collisions, does not apply here. Conservation laws cannot be applied to single balls since they are affected by every other ball at all times. As far as conservation principles go, a non-physical mind could give an instantaneous boost to any of the particles, as long as it also gives compensatory boosts to other particles at the same time.

Suppose, for example, that the "pineal gland," to use Papineau's idiom, is a galaxy of such point-particles. Suppose further that four particles in it, A, B, C and D, are heading exactly north, south, west and east, respectively. The mind could, for example, boost A heading north (changing its speed but not the direction of its movement), at the same time giving an equal amount of extra speed to B travelling south to preserve the system's overall momentum, while also slowing down C and D moving east and west respectively, to counterbalance the extra kinetic energy A and B are receiving, and so preserving both the overall energy and momentum of the system of particles that make up the pineal gland. If A and B are connected by a north-south line, C and B by a west-east line, which is to say that both pairs of particles move along the line that connects them, such a redistribution of energy and momentum would leave the overall angular momentum of the system untouched, too. So such an operation would be in conformity with all three of the relevant conservation principles (and the mind doing it could well act spontaneously). Such a mental intervention is possible, as far as conservation principles go, because conservation principles are not concerned with individual particles, since individual particles are never isolated from the rest of the particles in this world, they are only concerned with the aggregates, over the whole world, of some of their properties. The number of this system's degrees of freedom easily outruns the number of independent constraints, expressed in the form of independent equations, that these conservation principles entail to constrain the evolution of this system for any n bigger than 2.

The boosts given by the immaterial mind to these particles had to be instantaneous, though. Otherwise every change in a particle's speed would be accompanied with a change in its position, which would lead to a change in the system's potential energy. In the previous subsections we always considered changes that physical forces could bring about, and argued that there no reason to think that if an effect of an external physical cause is consistent with a conservation principle (applied to the system affected), so is the case when the external physical cause is replaced by a non-physical one. Instantaneous boosts cannot be produced by a physical force, because that would mean infinite acceleration, which in turn would require an infinite force. However, supposing that an appropriate symmetry obtains in the n-particle system, with the four particles in it which the mind accelerates, we could easily construe an example in which the changes of the four particles' velocities are accompanied by symmetrical changes in their positions that keep the potential energy of the system constant, so the instantaneousness of the boosts can be dropped.

Of course, the full classical dynamical description of this world is deterministic, so, if the classical description was the last word about the evolution of this world, it would not allow for law-conforming external influence. It is worth to pause to look into this classical dynamical description. It is of interest to our topic because, when explaining how the conservation of energy became a central tenet of modern science, Papineau rightly points to a post-Newtonian development in classical mechanics, known as "rational mechanics," which gave energy in a central role in the conceptual framework of classical mechanics it did not possess previously under more narrowly Newtonian terms. Papineau writes,

> Through the eighteenth and early nineteenth centuries a number of mathematician-physicists, among the most important of whom were Jean d'Alembert (1717-83), Joseph Louis Lagrange (1736-1813), the Marquis de Laplace (1749-1827) and William Hamilton (1805-65), developed a series of mathematical frameworks designed to simplify the analysis of the motion of interacting particles. […]
>
> These mathematical developments also implied that, under certain conditions, the sum of kinetic energy and potential energy remains constant. Roughly, when all forces involved are independent of the velocities of the interacting particles and of the time (let us call forces of these kinds "conservative"), then the sum of actual kinetic energy (measured by $1/2mv^2$) plus the potential to generate more such energy (often called the "tensions" of the system) is conserved: when the particles slow down, this builds up "tensions", and, if those "tensions" are expended, the particles will speed up again. (Papineau 2001.)

This, and everything else Papineau writes about rational mechanics in the article, is of course true. But it may give us the impression that in rational mechanics the full classical mechanical description of systems of particles held together by conservative forces is somehow traced back to the conservation of energy, conceived as the conservation of kinetic plus potential energy. And this is not the case.

So I propose to briefly review the classical dynamical description of our n-particle world on the basis of Hamilton's principle to see what exactly the dynamical description has and has not to do with energy and the conservation principles.

According to Hamilton's principle, there is a function of the locations and velocities of the particles, and of time:

$$L(\boldsymbol{r_1}, \boldsymbol{r_2}, \dots, \boldsymbol{r_n}, \boldsymbol{v_1}, \boldsymbol{v_2}, \dots, \boldsymbol{v_n}, t),$$

called the Lagrangian of the system, such that, between two fixed points in the evolution of the system, characterized by the values of the coordinates of the particles (that is, the full configuration of the system) at two different times, t_1 and t_2, the true path of the system's evolution (in the abstract 3n-dismensional space in which every point represents a possible configuration of the system) is the one for which the integral of it over time

$$S = \int_{t_1}^{t_2} L \, dt$$

is stationary, that is, its variation is zero. (If S takes its minimum for the true path, it is one way for it to be stationary. Most often, although not always, for the true path, S is minimal. This is why Hamilton's principle is often called, somewhat inexactly, the principle of least action, action being S, the integral of the Lagrangian over time.)

This variation principle, in turn, entails the differential equations that determine the motion of the particles.

So, to obtain a dynamical description for our world of interacting point-particles, we need to determine its Lagrangian. By hypothesis, in this world the interaction between the particles depend only on the location of the particles. Such an interaction can be captured by adding to the Lagrangian of a non-interacting n-particle system a term that is a function of the location of the particles. So we should first determine the Lagrangian of an aggregate of n non-interacting particles.

In general, the Lagrangian is a function of the locations and the velocities of the particles involved, and of time. Considerations about the homogeneity and isotropy of space, and about the homogeneity of time, entail that the Lagrangian of a single free point particle cannot depend explicitly on its location ($\boldsymbol{r}$), and the direction of its velocity vector ($\boldsymbol{v}$), nor can it depend explicitly on time (t). So it will depend on the absolute value of the velocity vector (v). From the Galilean relativity principle we can derive that it should be proportionate with the square of the absolute value of the velocity vector (for the details of the derivation, here and in the rest of this section, see e.g. the first volume of Landau and Lifshitz's Course of Theoretical Physics (1976)). Using the additivity of the Lagrangian, the Lagrangian of a non-interacting n-particle system would be

$$L = \sum_{i=1}^{n} \frac{m_i v_i^2}{2} \quad \text{(where the index i runs through the n particles.)}$$

With the addition of a term expressing the interaction between the particles, which depends only on their location: $-U\,(\boldsymbol{r_1}, \boldsymbol{r_2}, \dots, \boldsymbol{r_n})$, the Lagrangian of the interacting n-particle system takes the form

$$L = \sum_{i=1}^{n} \frac{m_i v_i^2}{2} - U\,(\boldsymbol{r_1}, \boldsymbol{r_2}, \dots, \boldsymbol{r_n})$$

Now according to Hamilton's principle, between two times, t_1 and t_2, the true path of the system's evolution is the one for which the integral

$$S = \int_{t_1}^{t_2} \sum_{i=1}^{n} \frac{m_i v_i^2}{2} - U\,(\boldsymbol{r_1}, \boldsymbol{r_2}, \dots, \boldsymbol{r_n})\, dt$$

is stationary, that is to say, for which the variation of S is zero. From this, the differential equation determining the movement of the i-th particle can be obtained as:

$$m_i \frac{d\boldsymbol{v}_i}{dt} = -\frac{\partial U}{\partial \boldsymbol{r}_i}$$

which means that the interaction between the particles accelerates each particle in the direction of the negative gradient of U. With the substitution of

$$\boldsymbol{F_i} = -\frac{\partial U}{\partial \boldsymbol{r}_i}$$

this result is takes the form of the well-known Newtonian equation of motion for the i-th particle.

This is a system of differential equations for the time-dependence of the coordinates, which, if the initial coordinates and velocities of the particles, and the U function characterizing the interaction between them are known, has a unique solution. (This entails that, classically, the evolution of this system is deterministic, which, in turn, entails that this system is closed to law-conforming external influence.)

The identity of the so far unspecified U function can be discerned from the consideration that, because of the homogeneity of time, the Lagrangian of a closed system cannot depend explicitly on time. From this it can be derived that

$$E = \sum_{i=1}^{n} \frac{m_i v_i^2}{2} + U\,(\boldsymbol{r_1}, \boldsymbol{r_2}, \dots, \boldsymbol{r_n})$$

is constant throughout the evolution of the system.

$$K = \sum_{i=1}^{n} \frac{m_i v_i^2}{2}$$

in this formula is the total kinetic energy of the n-particle system. The addition of U, a function of the location of the particles, to it, yields an energy-dimensioned quantity that is conserved.

$$E = K + U = const.$$

So the U function is the potential energy of the system arising from the interaction between the particles, which is a function of how the particles are located, and E is the total, that is, kinetic plus potential, energy of the n-particle system, which is preserved throughout the evolution of the system.

This reveals precisely what Hamilton's principle says about the energy of the n-particle system. In effect, it says that *the path of the system in the configuration space is the one along which the average difference between the system's kinetic and potential energy, $K - U$, is minimal (more precisely: stationary)*. This principle entails the evolution of the system unambiguously, so it entails that the evolution of the system is deterministic, and thus closed to law-conforming external influence. This principle has everything to do with energy. But it is not equivalent to, or entailed by, the conservation of energy. It requires, and has, independent empirical justification.

Similar considerations about the homogeneity and isotropy of space entail the conservation of the overall linear momentum and the overall angular momentum of the n-particle system. *Relative to Hamilton's principle* (inserting the correct form of the Lagrangian), the basic symmetries of physics and the conservation principles are equivalent. It doesn't mean, however, that Hamilton's principle, determining the evolution of the system, is entailed by the conservation principles.

To sum up the moral of the section, conservation principles could entail closure with respect to law-conforming external influence only if, at all times, there was only one way for physical systems to continue their evolution consistently with the conjunction of the conservation principles and their state at the given time. In other words, only if conservation principles entailed determinism. However, for systems above some minimal level of complexity, conservation principles are insufficient to entail determinism. Proponents of physical closure need to look into the full dynamical description of these systems to see if they are deterministic.

7 The status and significance of quantum indeterminacy

In this section, we look into the question whether the dynamics of physical systems is likely to be deterministic, as far as out best current theories go, and what if it is not. To keep track of the agenda that was outlined earlier, in this section we will discuss whether there are sound arguments for physical closure with respect to law-conforming external influence of the following sorts:

(D2) The (correct interpretation of the) dynamical laws of physics is deterministic, and this entails physical closure with respect to law-conforming external influence;

(D3) In case the correct interpretation of the dynamical laws is indeterministic, but the dynamical laws are probabilistic, this entails physical closure with respect to law-conforming external influence all the same;

and

(D4) The dynamics of neural states is deterministic, and this entails closure with respect to law-conforming non-physical mental influence.

In case it turns out that there is uncertainty about key premises to such arguments that only empirical science can supply, we will also look into what the prospects are for getting a clearer picture any time soon.

A) Are the dynamical laws (or the correct interpretation thereof) deterministic?

There can be little doubt about the conditional that *if* the evolution of the whole of the physical domain is, truly, not just approximatively and within a specific range, subject to deterministic dynamical laws, *then* the physical domain is closed with respect to law-conforming external influence, as it has already been acknowledged in the previous section. Since it would mean that, given a prior state, the entire evolution of the physical state of the world is entailed by the dynamical laws, if a non-physical factor could make a difference to it, it would be in violation of some dynamical law. The question is whether the antecedent of this conditional is true.

Before quantum mechanics, determinism seemed likely to hold because it was entailed by the empirically successful classical formalism. The relevant question to ask about determinism was whether the formalism could cover all physical phenomena.

For large areas of physical phenomena, which at a time seemed to be almost the entire domain of physics with only a very marginal remainder, dynamical laws were identifiable to predict the evolution of the state of physical systems (closed systems, or systems whose interaction with the environment could be captured by imposing some mathematically clear-cut boundary condition) in a seemingly precise manner, relative to an earlier state. These laws can typically be expressed in the form of differential equations. The solutions to these differential equations are functions that express the time-dependence of the measurable properties that characterize a system's physical state. Taken together with the initial conditions, that is, with the known values of these variables at an earlier time, these differential equations have unique solutions (they are Cauchy problems). If the evolution of every aspect of the physical state of the universe is subject to such laws, this amounts to determinism. Some phenomena, however, seemed to resist description in such terms, and this led to the development of quantum mechanics.

In quantum mechanics a story, up to a point quite similar to the classical one, comes with a twist. For every quantum mechanical system, there is a dynamical law, and it drives, completely deterministically, the evolution of a mathematical object which is intimately related to the physical state of the system in question. This mathematical object is most often called the state-vector (or wave function), the name suggesting that it mathematically represents the physical state of the system concerned, a piece of objective reality. But, as it turns out, there are different ways to think about what the formalism of quantum mechanics means, and the meaning of the state-vector is not the same on every account. Probably the only entirely uncontroversial thing to say about the state-vector (apart its mathematical features) is that it can be used to calculate the probabilities of possible observations.

Soon after its inception it has become clear that quantum mechanics correctly predicts a vast array of physical phenomena that previously defied classical description. And not just that. It can also explain, in its own terms, why the classical description is adequate in a very good approximation in the range of phenomena where it works well. Quantum mechanics clearly outperforms classical physics empirically, and also contains it as a limit, so to speak, so there is no doubt that it represents a huge leap forward in our quest for the true laws of nature. When confronted with a metaphysical question, such as the question of physical determinism, or that of physical closure, we have much more reason to take our clues from quantum mechanics than from classical physics.

Quantum mechanics, however, is not very easy to take metaphysical clues from.

This problem arises from the fact that the mathematical formalism represents the physical state of quantum systems as (normalized) elements of a vectorspace (adopting the talk that the state-vector represents the physical state, which is true on most but not all interpretations of quantum mechanics). The whole vectorspace represents the totality of possible states. A vectorspace is a set equipped with addition and multiplication by number (in our case, complex number), which is closed under these operations. (In fact, there is more mathematical structure here. The vectorspace used to represent quantum states is also equipped with an inner (scalar) product. On the basis of an inner product, one can define a notion of orthogonality, and can introduce a metric. Once a metric is in place, it makes sense to talk about the convergence of a series, and completeness, which is the property that Cauchy series composed of elements of the vectorspace have their limits within the vectorspace. The conjunction of these mathematical features makes this vectorspace a Hilbert space.)

Choosing a vectorspace to represent the set of possible physical states bears the important consequence, the source of much of the puzzlement, that a linear combination (a weighted sum) of two possible states is also a possible state. Formally speaking,

$$\Psi_1, \Psi_2 \in \mathcal{H} \;\&\; c_1, c_1 \in \mathbb{C} \supset c_1\Psi_1 + c_2\Psi_2 \in \mathcal{H}$$

The possible outcomes of measurements are represented as solutions to eigenvalue equations of operators acting on this space:

$$\hat{\mathrm{O}}\Psi = \lambda\Psi$$

With every measurable property a linear operator is associated. The only possible outcomes of the measurement of a property are the eigenvalues of the operator associated with it.

Another crucial feature of the formalism is that we use operators to represent measurables whose eigenvectors form a linear basis in the Hilbert space of all possible states. It means is that any possible state-vector can be expanded as the linear combination of the eigenvectors of any operator that represents an observable property:

$\Psi = \sum_i c_i \Psi_i$, $\hat{O}\Phi_i = \lambda_i \Phi_i$., for any Ψ and Ô, where *i* is an index running through all the eigenvectors $\{\Phi_i\}$ of Ô.[2]

When the system enters the measurement process in an eigenstate of the operator associated with the property that is being measured, the outcome of the measurement is the eigenvalue corresponding to the eigenstate, with probability 1. In any other case, the system is in a "superposition" state with respect to the property we are measuring, meaning that its state is a linear combination of the eigenstates of the operator that represents the property. In such cases a measurement of the property in question can yield any of the eigenvalues of the operator that represents the property, and the probability that the i-th possible outcome (λ_i) will be obtained is determined by the weight with which the eigenstate corresponding to it figures in the linear combination:

$$P(\lambda_i) = |c_i|^2.$$

This is the Born rule.

Different operators representing different properties may well have different sets of eigenvectors, which is not a problem in itself, since a vectorspace can have an infinite number of different bases. It bears the consequence, however, that if the system is in an eigenstate with respect to one physical property, chances are that the very same state is a superposition state with respect to another property. This feature of the formalism is closely related to the famous uncertainty principle.

And here is the last characteristic of the formalism that is responsible for the interpretational problems. The dynamical law that drives the evolution of the state-vector is also linear. It means that it preserves superposition: if $\Psi_1(o)$ evolves into $\Psi_1(t)$, and $\Psi_2(o)$ evolves into $\Psi_2(t)$, then the superposition state $c_1\Psi_1(o) + c_2\Psi_2(o)$ evolves into $c_1\Psi_1(t) + c_2\Psi_2(t)$.

This is how we predict the probabilities of measurement outcomes, supposing we know the state-vector of a system at some earlier time: We suppose that from that time until the time when the measurement will be performed, it evolves deterministically driven by the dynamical law. So we can predict its state-vector at the time when it enters the measurement process. Then we apply the Born rule.

How do we know the state-vector of a system at any time? Where do we get the initial conditions for our predictions? It is assumed that when a system is being measured, its state-vector becomes (in the usual jargon, collapses into, or reduces to, or get projected to) the eigenvector corresponding to the eigenvalue that is obtained as the result of the measurement, and resumes its normal evolution driven by the dynamical law starting out from this state. This is the state that needs to be fed into the prediction of any further measurement as the initial condition. So a measurement is also a preparation of the system that is being measured.

This scheme of predicting observations seems to be making an irreducible reference to the concept of measurement, which thus appears to be a metatheoretical concept. Nature is assumed to behave differently when it is being measured from the way it does when it is left alone, that is to say, there are two very different dynamical processes posited, a smooth and deterministic evolution for normal situations, and an abrupt and stochastic change (a 'collapse') for measurements. But then, we should wonder, how does nature know that it is being measured and therefore the state-vector should

[2] For the sake of simplicity, a discrete and non-degenerate eigenvalue spectrum is assumed.

perform a collapse, rather than carrying on with the smooth evolution driven by the dynamical law? The theory is generally perfectly capable of handling the situation when two systems get entangled. They would be treated as a bigger, composite system, there would be a larger Hilbert space to represent the possible states of the composite system, the tensor product of the Hilbert spaces representing the possible states of the two subsystems.[3] In it, a state-vector would represent the composite system's state at any time, and the evolution of the state-vector would be driven by the dynamical law. Why can't it work the same way when one of the two systems is a measurement apparatus? A measuring device, after all, is just another physical system consisting of elementary particles, the paradigm objects of quantum mechanical descriptions. Admittedly, it must be a big one for us to be able to directly observe some aspects of its physical state, otherwise we couldn't use it as a measuring device. But there seems to be no intrinsic indication in the theory that it should be applicable to collections of particles only as long as they are "small." It seems also hard to identify a non-arbitrary threshold that could demarcate "big" from "small."

However, as it turns out, applying the quantum mechanical description to the composite system composed of the measured system and the measuring apparatus, but just the unitary part of it, the evolution driven by the dynamical law, without the collapse, would lead to an absurdity. A measuring device is a macroscopic system that engages with the measured system in an initial state we may call the "ready state," and then, while the composite system evolves, its state evolves into one from which we can draw an inference with respect to what the state of the measured system was when it got entangled with the measuring device. Ideally, if the measured system gets entangled with the measurement apparatus when it is in an eigenstate of the operator representing the property that is being measured, at the end of the measurement process it will be in the same eigenstate, while the measuring device will evolve into a state which would amount to something like a pointer pointing to the corresponding eigenvalue on a scale or equivalent. Denoting the i-th eigenstate of the measured system with respect to the property that is being measured with Φ_i, the "ready" state of the apparatus with A_R, the end-state of the apparatus when its pointer points to the i-th eigenvalue A_i, the unitary evolution of the composite system could be written as:

$$\Phi_i \otimes A_R \rightarrow \Phi_i \otimes A_i$$

The same can be told of any other eigenstate. Now what if the measured system gets entangled with the measurement apparatus in a superposition state, that is, in a linear combination of the eigenstates? The linearity of the dynamical law entails that the measurement apparatus will end up in a linear combination of the end-states considered earlier:

$$\sum_i \Phi_i \otimes A_R \rightarrow \sum_i \Phi_i \otimes A_R$$

In which the pointer points where? A linear combination of different pointer positions makes no sense. We never see a measuring device with its pointer wiggling franticly whenever it measures a system in a superposition state. It suggests that there must be more to a measurement process than just the whole entangled system evolving deterministically, driven by the dynamical law. Apparently, to have a definite measurement outcome we need a collapse. But then, what triggers the collapse?[4]

---

[3] A tensor product of two vectorspaces can be thought of simply as an operation that merges them. Supposing we start off with an n-dimensional, and an m-dimensional vectorspace, the tensor product will be an m+n-dimensional vectorspace, and both original vectorspaces will be a subspace of it. The tensor product of an n-dimensional vector from the first vectorspace and an m-dimensional vector from the second will be an n+m-dimensional vector. If this vector is projected onto the two subspaces of the tensor product space that correspond to the two original vectorspaces, we get back the two original vectors.

[4] This consideration involved a simplified and idealized picture of the measurement process, but the resulting problem that the linear dynamical law would make the measuring apparatus evolve into a superposition of

This is the measurement problem. An equally appropriate name for it, preferred by some physicists and philosophers of physics, is “the macro-objectification problem.” The two names express the two sides of the same coin. It seems that either we somehow justify the special status of measurement and the associated duality of the dynamics, or we are at loss accounting for how a consistently superposition-free observed reality arises from the unitary linear dynamics which admits of, and preserves, superpositions, unless of course we assume that something was left out of the quantum mechanical description that accounts for that. Yet another name for the problem is “the quantum reality problem.” Although quantum mechanics predicts precisely and reliability what we are to observe, it is difficult to get a coherent ontological idea of what it is really that we observe.

The physical and metaphysical theories that have been developed to answer the measurement problem are the “approaches” to, or “interpretations” of, quantum mechanics. There are quite a number of them. Some of them command adherence in a significant portion of the scientific community, but each involves elements that other parts of the scientific community find hard to accept. As it happens, they diverge on the issue of determinism. The predictions of the formalism are of course generally probabilistic, and on some interpretations it reflects objective indeterminism in the evolution of physical reality. But on others, the indeterminism is merely epistemic. In order to get a grip on where we stand vis-à-vis the question of determinism, we need to survey at least the main (most popular) approaches to quantum mechanics, even if very briefly, and we need to assess the reasons, if any, to make preferences between them.

One possible attitude in the face of this puzzlement is of course to take the formalism merely instrumentally. Famously, this was how Niels Bohr thought about the theory:

> In our description of nature the purpose is not to disclose the real essence of phenomena but only to track down as far as possible relations between the multifold aspects of our experience. (Bohr 1961: 18)

Paul Dirac insisted that the theory is about our knowledge of a physical system, lacking any ontological import, and that the collapse of the superposed state-vector into an eigenvector represents our knowledge of the system becoming more precise in a certain respect (see Hendry 1984). This restriction made it into to a concordance achieved by a group including Bohr, Dirac, Heisenberg, Pauli and Born at the 1927 Solvay conference, which became the core of what is known as the “Copenhagen interpretation” of quantum mechanics. This name, however, is to some extent improper, as this way of thinking about quantum mechanics is more like self-restraint being exercised in the face of the difficulty of making an ontological sense of the formalism, abstinence from interpreting the theory rather than an interpretation, so, rather aptly, it is sometimes called the “shut-up-and-calculate interpretation” of quantum mechanics. The instrumentalist attitude is widespread in the scientific community until the present day, especially among practicing physicists.

Since our question of interest is what to make of the probabilistic character of the predictions of quantum mechanics, i.e. whether we should see it as a manifestation of the objectively indeterministic nature of the evolution of physical reality, or some other way, maybe reflecting an indeterminacy that is merely epistemic, it is interesting to ask ourselves how one leaning towards instrumentalism about quantum mechanics is likely to answer this question. On such grounds, we are certainly blocked from taking quantum indeterminacy at face value. Dirac’s view that the state-vector is about our knowledge of the physical system concerned seems to imply that we should see quantum indeterminacy as an indeterminacy in the evolution of our knowledge about the system under consideration. However, it is not a “merely epistemic” indeterminacy in the usual sense. It is assumed, on the Copenhagen view, that there are no conceivable further revelations about the physical state of the systems concerned that would make the predictions of experiences deterministic. The state-vector is assumed to embody everything that there is to know about physical systems. The adjective

---

definite-reading states is not dependent on the simplifying and idealizing assumptions. See e.g. Section 3 in Ghirardi 2009.

"conceivable," however, in the sentence before the last, would be a correlatum of *our* conceptual apparatus with which we attempt to grasp physical reality, and the Copenhagen orthodoxy would involve pessimistic comments about the adequacy of this conceptual apparatus for this purpose. One thing is sure though. Since the quantum mechanical description is empirically superior to the classical description, and since it can explain in its own terms why the classical description works as a very good approximation in the situations when it does, the determinism of the classical formalism cannot be regarded as evidence for the metaphysical thesis of physical determinism on the Copenhagen view (provided that such theses are regarded to make sense at all).

Einstein, for one, was thoroughly dissatisfied by this metaphysical self-restraint of the Copenhagen orthodoxy. He wrote,

> What I dislike in this kind of argumentation is the basic positivistic attitude, which from my view is untenable, and which seems to me to come to the same thing as Berkeley's principle, esse est percipi (Einstein 1951: 669).

But it is one thing to say that the theory admits of no intuitive and readily acceptable ontological interpretation, and it is another thing to say that it admits of no ontological interpretation at all. Metaphysical theories, even if extravagant in some respect or other, can be supplied to solve the measurement problem, either by embracing and justifying the collapse of the superposed quantum state, or by explaining away the need for a collapse, that is to say, by explaining away the apparent contradiction between superposition at the level of the quantum state, on the one hand, and the definiteness of observable phenomena, on the other. As a general rule, on the interpretations that embrace and explain collapse, the evolution of fundamental physical reality involves objective indeterminacy, whereas on those that do away with collapse, which then would leave us with the unitary evolution of the state-vector driven by the Schrödinger equation (and possibly some extra structure but no collapse), the probabilistic character of the observed phenomena is explained by an indeterminacy that is merely epistemic. There is an exception to this rule, though: there is no collapse on the consistent histories interpretation, yet the whole evolution of the quantum state is thought to be stochastic (Griffith's version).

One metaphysical theory to explain collapse came very early on, and can be viewed to a certain extent as an extension of the Copenhagen view. The Copenhagen interpretation already involves a retreat from the view that physics reports of a mind-independent reality, since it insists that all physics reports of is statistical expectations regarding future experiences, justifiable on the ground of earlier experiences and the physical laws. So it involves an irreducible reference to experience. John von Neumann (1955) and Eugene Wigner (1967, 1997) proposed an explicitly dualist interactionist metaphysical theory, according to which quantum mechanics reports of the interaction between physical reality and consciousness. This theory solves the measurement problem by assuming that it is the presence or absence of consciousness that distinguishes between measurement situations in which the quantum state undergoes a collapse, and non-measurement situations in which the quantum state evolves smoothly, driven by the dynamical law. Since experience is generally superposition-free, merely looking at things, or other cases of perception when we are bound to experience definite macroscopic phenomena, constitutes a measurement situation in this sense. The physical brain becomes entangled with the rest of the physical world, and when the mind turns its attention to a certain aspect of the state of the physical world, it collapses the quantum state of this entangled system into one of the eigenstates that correspond to the aspect of the world currently in the focus of the mind's attention. So the non-physical mind both selects a basis according to which the quantum state of the entangled composite system of the brain and the rest of the physical world will be expanded as a linear combination of eigenstates, and collapses the state. The collapse, in turn, gives rise to a definite experience in the mind. This is how the mind contributes to the evolution of the physical world, and the physical world contributes to the course of the experiences that arise in the mind. This is much like Cartesian dualism, except in this variety we are not left completely in the dark about how the interaction between the mind and the physical world is taking place. Sometimes this interpretation is presented as a relic of the early days of quantum mechanics, but this is not quite appropriate. The view may not be very popular nowadays, but it has always had advocates since the early days up to the present. A particularly forceful contemporary advocate is Henry Stapp (1993, 2007).

For many, this solution to the measurement problem is unacceptable precisely because of the metatheoretical status of the observer and the measurement situation. This is how David Bohm expressed this dissatisfaction:

> If the quantum theory is to be able to provide a complete description of everything that can happen in the world [...] it should also be able to describe the process of observation itself in terms of the wave functions of the observing apparatus and those of the system under observation. Furthermore, in principle, it ought to be able to describe the human investigator as he looks at the observing apparatus and learns what the results of the experiment are, this time in terms of the wave functions of the various atoms that make up the investigator, as well as those of the observing apparatus and the system under observation. In other words, the quantum theory could not be regarded as a complete logical system unless it contained within it a prescription in principle for how these problems were to be dealt with (Bohm 1952: 583).

It should be noted, first of all, that, contrary to Bohm's complaint, the way the consciousness-induced collapse theory was presented by von Neumann, the quantum mechanical description may very well encompass the particles that make up the measurement apparatus and the observer (due to the "movability of the von Neumann cut"). It falls short of presenting the quantum mechanical description as "a complete description of everything that can happen in the world" because it posits irreducibly non-physical mental events, and the physical description is incomplete regarding the *physical part* of the world (the atoms of the measurement apparatus and the observer included) because the theory posits interaction between the non-physical mind and the physical world. To require of a physical description that it encompasses everything that can happen in the world, is assuming physicalism. To require of a physical description that it accounts for every physical event without remainder, is assuming physical closure. In the context of trying to find out whether physical theory entails closure, a proponent of physical closure cannot regard these as valid criteria for the respectability of a physical theory without begging the question against the interactionist. It is sometimes claimed by advocates of the closure thesis that it is a working hypothesis of physics that the physical description should be complete exactly in the sense it is demanded by Bohm in this passage, and they would certainly say that this passage from Bohm stands witness to it. Well, then the work of von Neumann, Wigner, Heisenberg and others testifies with equal relevance that the completeness of physical description is not adopted as a working hypothesis by every physicist.

This is not to say, however, that it is impossible to account for the collapse of superposition states in a way that dispenses with the metatheoretical status of measurement and observers. What one has to do is to modify the dynamical law, so that the dynamical law itself can account for why collapse should take place in certain situations and not in others. Of course, it results in something other than a metaphysical interpretation of quantum mechanics. It amounts to a rival physical theory with a different formalism, and therefore, in principle, with different testable predictions in at least some range of phenomena.

This approach took shape most prominently in the work of Giancarlo Ghirardi, Alberto Rimini, Tullio Weber, and Philip Pearl (see e.g. Pearl (1976, 1979), Ghirardi, Rimini, and Weber (1986), and Ghirardi, Pearle, and Rimini (1990)), who proposed models that modify the dynamical law by adding stochastic and non-linear terms to the original equation of the standard theory to do precisely this job. Such models have two problems to solve, which are termed "the preferred basis problem" and "the trigger problem." The former arises from the fact that the Hilbert space in which state-vectors dwell admits of infinitely many linear bases, so any state can be represented as the superposition of eigenstates in different ways. So there should be an explanation why the spontaneous collapse of the state-vector should bring about the actualization of one of a particular set of mutually exclusive potentialities rather than one of any other set of such potentialities. The latter is the problem of proposing a mechanism responsible for the collapse that is sensitive to something like the "size" of the system, disposing of macro, but not of micro, superpositions. The GRWP model, named after the initials of the surnames of the authors just referenced, reasonably chooses the basis so that macro objects could always have definite locations, and proposes non-linear and stochastic terms to be added to the dynamical law which are sensitive to the number of particles involved, or, in a subsequent version, to the average particle number within an appropriate volume, thus, at bottom, posits a mechanism whose effect is negligible for microscopic systems, but highly relevant for macroscopic ones. The result is a unified dynamical theory, which accounts for microscopic systems the same way as the standard theory does, for micro-macro interactions such as measurements without the difficulties that arise if we assume the interaction between the measurement apparatus and the measured system to be governed by a linear dynamical equation, and for the classical behaviour of macroscopic objects. In micro-macro interactions this unified mechanics leads to the non-linear and stochastic collapse of the state-vector.

Given that the collapse of the superposition state is a stochastic process on this theory – and Nicolas Gisin (1989) offered an argument to the effect that adding nonlinearity without stochasticity to the dynamical law of the standard theory would be unacceptable because it would make superluminal signalling possible –, the view involves objective indeterminacy. Even though it does not involve explicit dualist interactionism as the von Neumann—Wigner interpretation does, it is hospitable to the kind of interactionism envisaged e.g. by John Eccles (1994) in his book *How the Self Controls its Brain* (see also Eccles 1990, and Popper and Eccles 1977). This sort of interactionism would consist essentially in the mind operating within a room for manoeuvre created in the physical causal order by indeterminacy, controlling, or at least affecting, which of the physically possible outcomes should occur in quantum mechanical processes taking place in the neural system, without contradicting physical laws. Whether such a suggestion can be made consistently with both neuroscience and the Born rule is the subject of the next two subsections.

This sort of interactionism is of course different from the kind suggested by von Neumann, Wigner and others, according to which the work of the mind would be selecting a basis and inducing the collapse. To distinguish them, they will go by different names in the remainder of this paper. Interactionism via the mind's causal role conceived as base selection and inducing the collapse of the state-vector will be called *von Neumann interactionism*, while interactionism via the mind's causal contribution conceived as controlling, or affecting, the result of the collapse will be called *Eccles interactionism*. The two kinds are by no means mutually exclusive: the von Neumann—Wigner interpretation requires one and may accommodate the other.

There is no such room for interactionism on theories that explain away, rather than justify, the collapse of the quantum state, and leave us with the unitary deterministic evolution of the quantum state. There are two strategies to do that, one is to assume that the state-vector does not encompass the whole truth about a quantum mechanical system, the other is to let go of the uniqueness of the observed outcomes of quantum processes.

One theory of the former kind started its career very early on, with the work of Louis de Broglie (1928). It soon became abandoned as the influence of the Copenhagen interpretation grew, but the abandoned idea has been taken up, or reinvented, by David Bohm (1952), resulting in a theory which gives a strikingly simple solution to the macro-objectification problem. The key to this solution is taking one particular (natural) mathematical representation of the quantum state at face value. This representation represents the quantum state with a wave-function. (A wave-function is a complex periodical function of the spatiotemporal co-ordinates. It is a possible representation of the state-vector because operations such as addition, multiplication by number, an inner product, and a metric based on the inner product can be introduced on the set of all possible wave-functions, making it a Hilbert space of the required kind.)

Particle-wave dualism was one of the problems that puzzled physicists greatly since the advent of quantum mechanics. The nature of the puzzlement is exemplified by the double-slit experiment.

When an electron leaves a mark on a screen that it reaches by going through either or both of two slits of a wall separating the source from the screen, it behaves like a small particle having a definite position. However, when a sufficiently large number of electrons are sent through the slits, the marks they leave on the screen build up an interference pattern that is characteristic of a wave coming through both slits at the same time, the two parts of it interfering with each other in the space between the slits and the screen. This is so *even if only one electron is emitted at one time*, and a subsequent electron is emitted only when the previous one has already reached the screen. The solution to this puzzle that de Broglie suggested was that there are in fact two entities, associated with each other: an electron is a combination of a wave and a particle. The particle is guided by the wave. The wave is the quantum mechanical wave-function that evolves according to the dynamical law, the Schrödinger equation. The particle is a classical entity, traversing along a definite deterministic trajectory. Regarding its choice between the two slits, and its subsequent trajectory from the slit to the screen, the movement of each electron (each particle) is determined by its initial position and the "quantum potential," analogous to potentials used in classical dynamics, derived from the phase of the wave-function. The wave goes through both slits, and interferes with itself subsequently. The extra information about the electron, not contained in the wave-function is its (the particle's) initial position, which we don't know. If it is assumed that its probability-distribution is $|\psi|^2$, where ψ is the initial wave-function of the electron, then the de Broglie-Bohm trajectories of an ensemble of electrons give out the interference pattern seen in the experiment. Bohm generalized this idea to systems of many particles, in which case the ontology consists of the particles and a single wave, the wave-function of the quantum mechanical many-particle problem. The wave evolves according to the Schrödinger equation that accounts for the interactions within the system, and it guides all the particles. The theory proved empirically correct in all non-relativistic experimental situations, which is little surprise, considering that it was designed precisely to reproduce the predictions of standard quantum mechanics.

The "macro-objectification problem" simply does not arise, given that all particles have definite positions at all times. The wave-function evolves always deterministically, driven by the Schrödinger equation. It never collapses. The theory, however, has to account for how it is that, in post-measurement situations, systems continue their evolution as if their wave-function had collapsed into an eigenstate. The answer Bohm gave to this question is an early version of decoherence theory. Decoherence, in general, is an account of how interference is suppressed between different branches of a superposed wave-function in certain situations. In Bohmian mechanics the component of the superposed wave-function that corresponds to the actual outcome of a measurement becomes the only one that is relevant for the post-measurement evolution of the system because it is the only component which significantly differs from zero at the well-defined locus of the system in the configuration space, so it is the only component that effectively guides it. For all practical purposes, the complex post-measurement wave-function can be replaced with the eigenstate corresponding to the outcome of the measurement. This feature of Bohmian mechanics is called the "effective collapse" of the wave-function in measurement situations.

With the "hidden variables" that make up the actual configuration of the system, the probabilistic character of the formalism's predictions is represented as being grounded in an indeterminacy that is of the epistemic, rather than the objective, sort. There is no contradiction here with John Bell's finding about the impossibility of completing the quantum mechanical description with local hidden variables, since Bohm's theory is expressly non-local. The movement of any one member of a many-particle system manifestly depends on the positions of all the others, however far they might be, as long as they are entangled, and the effect of a change in the position of a distant particle on the velocity of a particle here is instantaneous. Action at a distance is the price to pay for this solution to the measurement problem and restoring determinism. This price is not taken lightly by most physicists. A degree of non-locality is, however, arguably involved in any interpretation except those of the Everettian family and consistent histories, and all, including the Bohmian interpretation, admit of a formally Lorentz invariant generalization (even if with a preferred frame of reference, or a preferred foliation of spacetime, in the Bohmian case), so it is far from obvious how strongly it counts against the Bohmian theory.

Everettian interpretations promise a solution to the measurement problem without positing either non-local hidden variables, or an objective collapse of the quantum state. The central idea of these interpretations, in Michael Lockwood's words, is the rejection of "the assumption that when a measurement is carried out, one of the possible outcomes occurs to the exclusion of all the others" (Lockwood 1996). It is a non-negotiable datum, however, that an observer always finds herself in an unambiguous conscious state involving the experience or memory that she has observed one of the possible outcomes. There is only one way to reconcile this datum with the rejection of the assumption that one possible outcome occurs to the exclusion of all others, and this is to assume, flying boldly in the face of common sense, the coexistence of multiple parallel and mutually inaccessible conscious observer states after measurement. In Everett's own words,

> with each succeeding observation (or interaction), the observer state "branches" into a number of different states. Each branch represents a different outcome of the measurement and the *corresponding* eigenstate for the object-system state. All branches exist simultaneously in the superposition after any given sequence of observations. The "trajectory" of the memory configuration of an observer performing a sequence of measurements is thus not a linear sequence of memory configurations, but a branching tree, with all possible outcomes existing simultaneously in a final superposition with various coefficients in the mathematical model (Everett 1957).

It is a further question whether or not this branching of the observer's mental state supervenes on a branching of the physical state of the observer's brain and everything that is entangled with it, ultimately the whole world. We end up with "many minds" (Albert and Loewer 1988, Donald 1990, 1995,[5] Lockwood 1996) or "many worlds" (DeWitt 1970, Graham 1973, Saunders 1993, 1995, Deutsch 1996, Wallace 2012) interpretations, depending on how we answer this question. Instead of explaining collapse, Everettians must explain decoherence, i.e. why it is that for the further evolution of each post-observation branch, every other branch becomes irrelevant. Decoherence is a sensible and fairly developed theory. It is not the whole solution, however. Just as spontaneous collapse theories, Everettian theories also have to solve the preferred basis problem, arising from the fact that a state-vector can be spelled out as a superposition of eigenstates in many different ways. Apparently, these problems are more pressing for the many worlds than the many minds interpretations, since on many minds interpretations the physical state doesn't have to branch. And finally, there is also the problem of making sense of the probabilities of possible outcomes in a framework which is deterministic and in which every possible outcome gets actualized. This is really pressing because quantum mechanics predicts probabilities, and if we cannot make sense of them, the empirical success of quantum mechanics, which makes its interpretation a worthy task to start with, disappears.

And finally, there is the consistent histories approach, first developed by Robert Griffiths (1984), and subsequently by Roland Omnés (1994, 1999), and largely independently, under a slightly different name (decoherent histories) but containing similar essential ideas, by Murray Gell-Mann and James Hartle (1990). Its formalism can be used for the purposes of different approaches to quantum mechanics. Notably, some versions of the Everettian interpretation construe the branches of the Everett multiverse as consistent histories (Saunders 1993). Griffiths, however, developed it into a full-blown single-world interpretation in its own right. This short outline follows the way he thinks about consistent histories.

This approach is considerably more difficult to give a brief and non-technical outline than the others. We may start by reciting that on this interpretation the "state-vector" is not a representation of the physical state of a quantum mechanical system. The "state" of system is just the state of affairs that it instantiates certain properties, and a property that a physical system instantiates is represented

[5] Donald's is a Berkeleyan idealist variety.

by a subspace of the Hilbert space associated with the system, or, we could equally say, by the projector that projects onto the subspace corresponding to the property. (A projector is an operator that equals to its square. Any vector in the subspace onto which it projects is its eigenvector with eigenvalue 1.) It doesn't have to be a pure state (a pure state would be one-dimensional subspace of the Hilbert space). Properties that a physical system instantiates can be coarser of finer grained.

Essential to this approach to quantum mechanics is the logic of the predicates that can be predicated of a quantum system. The negation of a quantum property is represented by the orthogonal complement of the subspace corresponding to the property, or with the projector $I - P$, where I is the identity operator, and P is the projector that projects onto the subspace corresponding to the property. Generally, there may be a vast number of vectors in the Hilbert space which are neither in the subspace corresponding to the property, nor in the subspace, its orthogonal complement, corresponding to its negation. Equally importantly, it does not always make sense to talk about the same system instantiating two different properties at the same time, or about the same system instantiating either one or the other. The conjunction or disjunction of two properties makes sense only if the projectors representing them commute. A system at a time can be characterized from different points of view, and not all perspectives can be combined. A *projective decomposition* of the identity operator, a collection of orthogonal projectors that sum to the identity, represents the possible states of a system from a single point of view. The orthogonality of the projectors in the projective decomposition guarantees that the corresponding properties are mutually exclusive, and that they sum up to the identity guarantees that one of them actually is instantiated by the system. The "event algebra," an event meaning that the system instantiates one of a certain set of mutually exclusive possible properties, represented by a particular projective decomposition, at a time, is called "a framework." Two frameworks are compatible only if they are represented by projectors such that any projector from one projective decomposition commutes with any projector from the other. Otherwise they are incompatible, meaning that the two projective decompositions represent two descriptions of the system that cannot be combined. We may choose any perspective to describe a quantum system, but not any two perspectives can be combined to get a "fuller picture." This is the "single framework rule."

> [C]hoosing a framework is something like choosing an inertial reference frame in special relativity. The choice is up to the physicist, and there is no law of nature, at least no law belonging to relativity theory, that singles out one rather than another. Sometimes one choice is more convenient than another when discussing a particular problem; e.g., the reference frame in which the center of mass is at rest. The choice obviously does not have any influence upon the real world. But again there is a disanalogy: any argument worked out using one inertial frame can be worked out in another; the two descriptions can be mapped onto each other. This is not true for quantum frameworks: one must employ a framework (there may be several possibilities) in which the properties of interest can be described; they must lie in the event algebra of the corresponding PD [projective decomposition]. (Griffiths 2013.)

The evolution of a quantum system, in turn, is conceived as the system instantiating different properties at different times. Future (or past) states are not, in general, determined by the present state, they are only related to it by certain probabilities (which may be 0 or 1 in special cases, but that is the exception, rather than the rule). The unitary time evolution of the "state-vector" can be used to calculate conditional probabilities relating states of the same system at different times using the Born rule (the quotation marks indicate that, as it has already been noted, on this interpretation, the "state-vector" is not the representation of the physical state, it is generally just a tool to calculate the probabilities that the system will instantiate a property at a certain time). In order to assign probabilities to *histories* (the same system instantiating certain properties at certain times), one has to construct an appropriate sample space. For this we use a tensor product of copies of the system's Hilbert space, its subspaces representing the system's possible properties at different successive times.

$$\mathcal{H} = \mathcal{H}_0 \odot \mathcal{H}_1 \odot \dots \mathcal{H}_f$$

can be used for the description of the properties a system instantiates at successive times, t_0, t_1 … t_f. ($\mathcal{H}_0$, $\mathcal{H}_1$,… $\mathcal{H}_f$ are the very same Hilbert space, the lower index indicates different times at which the system's state is described. The symbol $\odot$ is used instead of the usual $\otimes$ only to signal that a time-sequence is being considered. Mathematically, the two symbols equally mean tensor product.) The possible histories of the system are represented in the tensor product Hilbert space $\mathcal{H}$. A history then is defined as the corresponding tensor product of projectors that project onto the subspace of $\mathcal{H}_i$ that represents the property the system instantiates at t_i:

$$F_0 \odot F_1 \odot \dots F_f$$

(F_i is the projector that projects onto the subspace that represent the property that the system instantiates at t_i.)

The collection of all possible histories of the system can be represented as

$$\{Y^\alpha\}$$

$$Y^\alpha = F_0^\alpha \odot F_1^\alpha \odot F_f^\alpha$$

called the "sample space," satisfying the condition that

$$\sum_\alpha Y^\alpha = I$$

I being the identity operator of the histories Hilbert space $\mathcal{H}$. Distinct histories belonging to the same sample space are mutually exclusive, one and only one of them is the history that actually obtains.

Just in the case of describing a system at a single time, describing a system at multiple times can be done from many different perspectives, and not any two of them can be combined. The single framework rule applies: incompatible sample spaces cannot be combined. When only two times are considered, probabilities can be straightforwardly assigned to histories of a closed quantum system using the Born rule, and the operator of the unitary evolution of the "state-vector" (again, just a mathematical tool to calculate probabilities) between the two times. This assignment of probabilities can be generalized to any number of discrete but arbitrarily chosen times respecting a consistency condition (which will not be discussed here).

This stochastic dynamical description applies to quantum mechanical systems at all times, and irrespective of whether one part of a composite system is a measuring device. If it is, then, the evolution of the system can be described using a framework in which the different possible pointer positions, so to speak, of the measuring device make sense, and then assigning probabilities to possible histories using the general rules. Since it is not the state of the composite system (the system instantiating physical properties) that evolves driven by a linear dynamical law, only a mathematical object that is used to calculate conditional probabilities of states obtaining at different times, the measurement problem does not arise.

The emerging picture is summed up by Griffiths as follows.

> The consistent ontology for quantum mechanics […] is realistic: the real world is "out there", not just a part of some observer's consciousness, and its structure is reflected in the mathematical theory constructed by physicists for describing it. But of course it differs from its classical predecessor in important ways, which can be conveniently summarized under two headings. First, the Hilbert space description of a system, in which its

properties are represented by subspaces, requires a new logic in the sense of a mode of reasoning about the world, with rules somewhat different from those familiar in classical physics, where ordinary propositional logic fits very comfortably onto an algebra of physical properties corresponding to subsets of the phase space. Second, quantum dynamics is intrinsically stochastic or probabilistic: probabilities are present in the basic axioms that apply without exception to all quantum processes, not just to those associated with some form of "measurement." (Griffiths 2013.)

The account is also perfectly local, and free of the striking ontological assumptions of the Everettian interpretations.

As far as I can tell, most physicists and philosophers of physics endorse a version of one or another of the approaches to quantum mechanics just sketched. None of them is without difficulties and implications that might strike us as implausible, but each broad option seems to be alive, and this is the main point for our present discussion. Since some of these approaches to quantum mechanics bear differently on determinism, the fact that they are at present pursued by significant portions of the scientific community would count decisively against anyone who would wish to claim that determinism is an established scientific fact, so it can be used as premise for a justification of the closure thesis, at least with respect to law-conforming external influence.

Since our topic extends to look into the prospect of the empirical justifiability of closure, even by research that has not yet been performed or completed, a quick note on the prospect that a choice between these different approaches to quantum mechanics will be made possible to make by yet to be obtained empirical evidence is in order.

The odds do not seem to be in our favour, since all of these approaches that are proper interpretations are designed to make sense of the same formalism, and consequently, have exactly the same testable predictions. Spontaneous collapse theory is an exception. It modifies the dynamical law to explain collapse, so it should be regarded not so much as an interpretation, but as a rival physical theory, testable against quantum mechanics. Giancarlo Ghirardi (2016) in his recently updated SEP entry on collapse theories has given an overview of the ongoing empirical research projects that aim to do precisely this. It seems that a body of empirical evidence to potentially confirm or disconfirm this particular approach to quantum mechanics can be expected to emerge real soon.

There doesn't seem to be much prospect, however, to differentially confirm or disconfirm the theories that leave the formalism that is responsible for the testable predictions untouched.

Except maybe by a desperate experiment Max Tegmark proposed a few years back to verify the Everettian interpretation. It has been termed "the quantum suicide experiment," and this is how Tegmark describes it:

> It requires quite a dedicated experimentalist, since it amounts to an iterated and faster version of Schrödinger's cat experiment – with you as the cat. The apparatus is a 'quantum gun' which each time its trigger is pulled measures the z-spin of a particle. It is connected to a machine gun that fires a single bullet if the result is 'down' and merely makes an audible click if the result is 'up' [...]. The experimenter first places a sand bag in front of the gun and tells her assistant to pull the trigger ten times. All [interpretations of quantum mechanics] predict that she will hear a seemingly random sequence of shots and duds such as 'bang-click-bang-bang-bang-click-click-bang-click-click'. She now instructs her assistant to pull the trigger ten more times and places her head in front of the barrel. This time the 'shut-up-and-calculate' recipe is inapplicable, since probabilities have no meaning for an observer in the dead state [...] and the [interpretations] will differ in their predictions. In interpretations where there is an explicit non-unitary collapse, she will be either dead or alive after the first trigger event, so she should expect to perceive perhaps a click or two (if she is moderately lucky), then 'game over', nothing at all. In the MWI [Everettian many worlds interpretation], on the other hand, the […] prediction is that [the experimenter] will hear 'click' with 100% certainty. When her assistant has completed his

> unenviable assignment, she will have heard ten clicks, and concluded that the collapse interpretations of quantum mechanics are ruled out to a confidence level of $1\text{-}0.5^n \sim 99.9\%$. If she wants to rule them out 'ten sigma', she need merely increase *n* by continuing the experiment a while longer. Occasionally, to verify that the apparatus is working, she can move her head away from the gun and suddenly hear it going off intermittently. (Tegmark 1997.)

It seems quite clear that the quantum suicide experiment could in principle verify the Everettian interpretation. It has, of course, never been performed. The reporter of New Scientist (issue 2113, 20 December 2007, p. 50) who interviewed Tegmark remarks that Tegmark suggested to him that "Perhaps I'll do the experiment—when I'm old and crazy." Even if he does, however, which I hope he won't, it will not be very instructive for most of our selves dwelling in parallel Everettian realities. Even if MWI is correct, 99.9% of our selves will read in the newspapers that he died trying to verify his favoured interpretation of quantum mechanics. He will be the only one who will not have the vast majority of his selves believing that the experiment ended tragically—for the simple reason that those copies of him will be no more.

There are of course considerations that might reasonably motivate an empirically undetermined choice between rival theories. Rival interpretations of quantum mechanics are no exception in this respect. Consistency with other physical theories is the weightiest of such considerations. In the case of quantum mechanics, the main concern is how it fits with relativity theory, an equally deep, beautiful and empirically successful achievement of modern physics. Although a Lorentz invariant generalization of quantum mechanics has been produced as early as the late 1920s, the difficulties of marrying quantum theory with general relativity are yet to overcome. This is a different problem, over and above the problem of giving a plausible ontological interpretation to the quantum mechanical formalism, but the two problems interfere with each other, since not all interpretations of the non-relativistic formalism are equally easy to generalize to the relativistic case in a way that fully respects the equivalence of inertial observers. But then, it is not entirely clear that it should rule decisively against an interpretation if it relies on a preferred frame, or a preferred foliation of spacetime, especially that Big Bang cosmological models of the spacetime of our universe seem to point to a preferred foliation. If our final aim is to assess the likelihood of determinism, discriminating between the different interpretations of quantum mechanics on the ground of their special relativistic generalizability would not be much help anyway, since there are both deterministic and indeterministic interpretations that pass this test with flying colours.

There are of course also softer principles of theory choice. However, unlike solid empirical evidence and consistency with other physical theories, these softer principles may compete on an equal footing with one's metaphysical preference for a theory that secures a physical description that is complete and closed, as we have seen in the passage cited earlier from Bohm, or, on the contrary, for a theory compatible with the idea that mental items that seem to resist a physicalist reduction, such as the subjective qualitative character of experience and the content of intentional states, may actually make a difference to the evolution of the physical state of the world. So, I think, when the issue at hand is the justification of the closure thesis, these softer principles of theory choice are of little use for us.

Perhaps the difficulty of giving a plausible ontological interpretation to its formalism and its tension with general relativity signal that quantum mechanics is not the last word on the dynamics of the physical state of the world, even if only forces other than gravity are concerned. Historically, it is not unusual even for empirically highly successful physical theories to turn out to be valid only approximatively, and to apply only in a limited domain, as a limit of a more general theory one step closer to the truth. Perhaps the same fate awaits quantum mechanics.

We are confronted here with the problem of trying to assess the justifiability of the closure thesis on the ground of an incomplete physics. We cannot do any better than look into the best theories we currently have. It may well be the case that one, or perhaps more, of the currently competing approaches to quantum mechanics contain hints that will lead us on towards an even better theory. But as far the present state of play is concerned, the question of determinism is wide open.

Having this clarified, it is time now to address the arguments of proponents of the closure thesis who believe that even if an indeterministic approach to quantum mechanics prevailed, that would be irrelevant. There are two views, in particular, that need to be assessed: the view (earlier labelled D3) that probabilistic laws entail closure just as much as deterministic laws do, and the other one (earlier labelled D4) that interactionism fails because determinism prevails at the neural level, even if the fundamental physical level turns out to be indeterministic. I assess these claims in Subsections B and C, respectively. Needless to say, we are still concerned with closure with respect to law-conforming external influence, even if it is not specifically mentioned.

B) IS CLOSURE ENTAILED BY PROBABILISTIC QUANTUM MECHANICS?

I have two arguments in mind, in particular, whose merits need to be assessed. The first of them is due to Ted Honderich, and its upshot is that combining indeterminism about quantum mechanics with interactionism, or at least with what we have earlier called Eccles interactionism, which is the only kind he considers, is incoherent.

Honderich (2002: 76) argues that Eccles interactionism is in fact a hidden variable interpretation of quantum mechanics, in which the self or originator, irreducible to the physical states and events taking place in the brain, plays the role of the hidden variable filling in the gaps in the quantum mechanical explanation for the course of physical events. Honderich thinks that an indeterminist-interactionist such as Eccles has, on the one hand, to embrace the completeness of quantum mechanics, thereby denying the possibility of hidden variables, in order to be a physical indeterminist, and then, on the other, adopt a hidden variables theory, to be in a position to deny that quantum mechanical events at the synapses are mere chance events.

This is not a good argument. To see this, we have to distinguish between two broad ways of thinking about non-physical mental influence in the course of physical events, say, at the synapses.

One possibility is to think of it as an extra force which cannot be traced back to the known physical forces, but which, nevertheless, is subject to laws, much like physical forces. In Section 6, we have discussed an argument by Papineau (2001) to the effect that this is the only way to think about of an external influence in the course of physical events which is consistent with the conservation of energy. We have seen that Papineau's argument rests on a doubtful premise which has no support whatever from physics, so we are not bound to think of the causal work of a non-physical mind this way. But it certainly is an option. On such a view, the mind seems to be just another kind of stuff which has its own laws for interacting with normal stuff. Either it is the case that the laws covering both physical and mental stuff, together with the initial conditions, entail the evolution of physical stuff, or, to the extent the evolution of physical stuff is left underdetermined by this full set of laws plus initial conditions, it is random, and maybe subject to probabilistic description.

An alternative way is to think that the mind is no stuff at all. This is to say that it is not in virtue of some non-physical properties that the mind instantiates, which subsume to psychophysical laws or are otherwise bound to influence matter, that the mind controls or influences its brain. Maybe there is no impersonal account for it. Maybe a person is a metaphysical primitive, and maybe a person has a metaphysically primitive power to select from physically possible alternative courses of brain events, or to make some of the alternatives more likely, and how a person choses to influence physical events is not explained by what she is and how she is at the time the influence takes place. If this exertion of influence is preceded by a consideration of reasons, the reasons do not cause the exertion of influence, either deterministically or probabilistically. Maybe the concept of "considering reasons" cannot be cashed out in an impersonal manner either. Exerting this influence always involves an element of genuine spontaneity.

This is all very rough and sketchy, but it will do as long as our goal is only to see if either of these ways of thinking about mental influence on the course of physical events in the brain, if it is imagined in the context of a quantum mechanical description, suffers from the confusion alleged by Honderich.

In the first case, if the mind is just another kind of stuff, interacting with physical stuff in a law-governed manner, nothing prevents us from extending the physical description to cover it, too. Then the subject matter of complete or extended physics will not only be what we previously thought it would be (roughly, matter), but it will be everything that is subject to strict laws of nature ('strict' is meant to be compatible with 'probabilistic'). Suppose now, that the relevant nomological description will turn out to be much like present-day quantum mechanics. Then, we may suppose, the configuration space in which the wave-function dwells would be added extra dimensions to cover for the mental degrees of freedom, and the Hamiltonian would be added extra terms to account for the so far neglected mind-matter interactions. But then this is not a case of mental influence taking place in a leeway in the physical causal order created by the probabilistic, rather than deterministic, character of the relevant laws. The quantum mechanical account for the (narrowly) physical subsystem of the composite (physical plus mental) system under consideration, which ignores the relevant mental forces, cannot be correct even probabilistically.

In the second case, on the other hand, the mental influence on the course of physical events we are considering is not a potential subject for any conceivable extension of the physical (nomological) description. It is irreparably external to it. The quantum mechanical account of the dynamics of the relevant system is then concerned only with the physical forces present, it is complete only in this restricted sense, and the interactionist does not claim otherwise. It is exactly the same as it would be in the counterfactual zombie situation when the relevant physical system is not inhabited by a non-physical mind. If the quantum mechanical description gives empirically correct predictions even in the actual situation when there is a mind to possibly interfere, it is either because the mind chooses not to, or because it interferes in subtle ways, bringing about only physical situations that could have obtained without its interference, and only without disrupting the probabilistic predictions of the quantum mechanical description. There is no confusion in it, unless the thought that the mind exerts an influence on the course of physical events without disrupting the Born rule is confused. So this is the problem I turn next.

It is quite customary to claim that quantum indeterminacy does not affect the problem of physical closure in any interesting way, since, if the course of physical events is subject to a probabilistic quantum mechanical account, an external influence would constitute a breach of the Born rule. For Papineau, the issue is only worth a footnote: for him, whether all physical occurrences, or *the chances of* all physical occurrences, are fully determined by prior physical history, makes no interesting difference as far as the admissibility of external influence is concerned. "I shall ignore this qualification in nearly all that follows, since it would only complicate the issues unnecessarily," Papineau declares early on in the Rise of Physicalism (note 2). On this point, he gets a rejoinder from Stapp, a dualist interactionist who likewise claims that Eccles style interactionism would "upset the logical coherence of the whole scheme" (Stapp 2007: 310), because it would disrupt the Born rule.

This dismissal is way too quick. The Born rule is about the probabilities of possible physical outcomes. To see if Eccles interactionism would indeed contradict the Born rule, we have to consider what it is that we really mean when we say that the probabilities of possible quantum events are determined by the Born rule.

It certainly has to do with the relative frequencies of outcomes obtained in a series of experiments performed with identically prepared quantum mechanical systems. One option is simply to identify the probabilities, predicted by the Born rule, with relative frequencies. Our initial shot to give a frequency interpretation to quantum mechanical probabilities could be to stipulate that the probability of outcome A is the relative frequency of outcome A in the reference class comprising all the outcomes of the experiments performed on identically prepared systems. Let us consider now a finite reference class consisting of the outcomes of a finite series of trials. It is easy to see that the relative frequency definition of probability is a nonstarter if the reference class is small. A fair coin will never land tails half of time if it is tossed only once. It may well land heads six times out of ten trials. The problem generalizes to any number of trials. We would certainly not call a coin unfair just because it landed 499 (instead of 500) times heads out of a thousand trials. We would expect the deviation from relative frequency .5 to be smaller, relative to the number of trials, if we make more trials, supposing that the coin is fair, but we cannot expect it to vanish completely, all the time. Of course, the same can be said also if the reference class consists of the outcomes of a quadrillion trials.

To make problems worse, we will never get a .5 relative frequency for a fair coin landing heads if it is tossed an odd number of times. Nor will we ever get an irrational number for a relative frequency. As long as we use a finite reference class, it will always be one integer over another, whereas in quantum mechanics, the Born rule often assigns irrational probabilities to possible outcomes.

For these reasons some frequentists proposed to use limiting relative frequencies in an infinite (and, therefore, hypothetical) sequence of trials (cf. e.g. von Mises 1957). This proposal has its own problems (see e.g. Hájek 2009), but at least it dodges these obvious ones which make finite frequentism useless for the purposes of quantum mechanics.

What is important for our discussion about this proposal is that there is absolutely no way for an immaterial mind to disrupt the Born rule, if probabilities are interpreted as limiting relative frequencies in hypothetical infinite sequences of trials, as long as it interferes with the outcome only a finite number of times. The change in relative frequency brought about by a singular interference gets smaller as the reference class gets bigger. The change in relative frequencies brought about by any finite number of interferences tends to zero as the size of the reference class tends to infinity.

Someone might want to object at this point that, if nonmaterial minds can interfere, we may presume that in a hypothetical infinite series of trials there would be an infinite number of such interferences. Well, there is plenty of room even for that, as long as the effects of these interferences cancel out statistically on the (infinitely) long run.

To sum up, the claim that Eccles interactionism would disrupt the Born rule seems very far from being obvious, if the probabilities that the Born rule assigns to quantum events are given a frequentist interpretation. Small actual reference classes will not do. In sufficiently big reference classes, the effect of a singular interference on relative frequencies is vanishingly small, and some deviation from the predicted relative frequencies should be allowed anyway. Limiting frequencies in hypothetical infinite reference classes have a much better job at interpreting quantum mechanical probabilities, but relative to them the effect of any finite number of interferences is zero. If we think that once the mind is allowed to interfere, we are bound to think it interferes an infinite number of times in a hypothetical infinite reference class, even that is fine, as long as the effect of the interference is statistically neutral. Admittedly, statistical neutrality would pose a serious constraint on Eccles interactionism, but it would not render it impossible.

It is time now to consider the other option, i.e. that instead of simply identifying probabilities with relative frequencies we think of physical probabilities as inherent tendencies of possibly differing strengths in a given type of physical situation to yield certain outcomes, and treat relative frequencies as imperfect evidence for probabilities so conceived. Provided that the conditions that define a type of physical situation are repeatable, and are in fact repeated a large number of times, these situations will yield certain types of physical outcomes with relative frequencies that approximate the respective probabilities.

Intuitively, this view, the propensity view of probability, sits well with standard quantum mechanics. Pictorially, when a quantum mechanical system enters a measurement process with its normalized state-vector pointing at a certain "direction," its orthogonal projections to the directions of the eigenvectors representing the possible measurement outcomes are of different lengths, and, according to the Born rule, these lengths determine the probabilities of these outcomes. This tells us how much the state-vector is already aligned with each of its options to reduce to in the process. It feels natural to interpret this as a representation of a set of inherent tendencies of differing strengths in the pre-measurement system to produce each of the possible outcomes.

Unlike the frequency view, which makes probabilities properties of mass phenomena (actual or hypothetical), the propensity view allows us to attribute probabilities to singular events – and one does find such attributions in quantum mechanics. The problem of irrational probabilities, which renders finite frequentism useless for quantum mechanics, simply does not arise, as relative frequencies are required only to approximate probabilities. So it seems there is much to say for a propensity view of probability in the context of quantum mechanics, which clearly was a motivation to propose the view for some of its proponents (cf. e.g. Popper 1957).

Let us suppose for the moment that the propensity view is correct; quantum mechanical probabilities are indeed intrinsic tendencies in the respective physical systems to produce certain outcomes. If a fundamentally non-physical mind selects one of the physically possible outcomes, it does not speak to the question what intrinsic tendencies there were in the physical system before the non-physical mind interfered. Such an extrinsic interference cannot disrupt, contradict or bias any natural law that assigns probabilities, as long as probabilities are interpreted as intrinsic tendencies in physical systems to produce possible outcomes.

To conclude, the claim that Eccles interactionism would disrupt the Born rule is very far from being obviously true. In fact, it is false unless we assume a) that a frequentist interpretation of probability is true, b) that if the mind is allowed to interfere at all, it would interfere an infinite number of times in a hypothetical infinite reference class, and c) that these interferences cannot cancel out statistically. Anyone claiming that this kind of interactionism is incompatible with the Born rule would be required to defend all three of these assumptions, and it would be a tough job.

And of course, Eccles interactionism is only one of the two ways an interactionist might want to take advantage of quantum indeterminacy. The other way, von Neumann interactionism, has nothing to do with the Born rule at all. As it was discussed in the previous section, on this view, the mind does not select (or increase the probability of) one of the possible outcomes in the indeterministic collapse of the state-vector into one of the eigenstates, rather, it poses the probing question, that is, it defines the set of possibilities, the set of eigenstates in terms of which it is relevant to expand the state-vector, and to one of which the state-vector is to reduce. On this view, quantum mechanics is not a theory of the evolution of the physical state of the world with causal gaps in it that allow for interference by the irreducibly mental, but is itself a theory of the interaction of mind and matter. On such a view, contradicting the Born rule, or Honderich's hidden-variable objection, are out of the question.

Admittedly, from the perspective of defending interactionism, von Neumann interactionism has clear disadvantages to Eccles interactionism. First of all, whereas Eccles interactionism is a possibility under any indeterministic interpretation of quantum mechanics, von Neumann interactionism is a feature of one specific interpretation, one which no longer has many adherents (although it does have some prominent proponents even today). Secondly, as David Chalmers remarks in *The Conscious Mind* (1996), "the kind of causal work consciousness performs here is quite different from the kind required for consciousness to play a role in directing behaviour." Indeed, structurally, it is as if in determining what we should do, our non-physical self would be allowed to determine only what question we ask of ourselves ('On which day of the week shall I go to holiday?', or 'Where shall I go for holiday?'), and then an answer would emerge to the selected question, but which of the possible answers, would be left a matter of pure chance. It does not amount to controlling our behaviour, at least, not in the usual sense, and, for Chalmers, in this respect it is not an interesting alternative to the epiphenomenalism that seems to follow from physical closure (combined with his argument for dualism). It is a sort of interactionism, though, so it is a possibility to consider for the purposes of our present discussion.

As a short digression, let us turn our attention back to the discussion of the meaning of physical probability in the context of Eccles interactionism. A few paragraphs above, we have established that on an indeterministic interpretation of quantum mechanics, and on a propensity interpretation of physical probability, interactionism Eccles style would not contradict the Born rule. The reason for that is that on a propensity interpretation physical probability is an intrinsic tendency of varying strengths in a physical system to produce certain possible outcomes (the strength of the tendency to produce a certain outcome being the probability of that outcome), and an extrinsic influence cannot come into conflict with any law that assigns to outcomes probabilities so conceived. It is important to note, and this leads us back to an earlier discussion in the Section 5, that giving a propensity interpretation to the probabilities that figure in probabilistic physical laws entails that, as predictors of the relative frequencies of possible outcomes, those laws should be regarded as ceteris paribus laws. (It also means that, if we stick to the terminology we introduced earlier, Eccles interactionism would be a case of overriding external influence (even though it does not contradict any law); so eventually we departed from discussing law-conforming external influence, in the sense defined earlier, and reobtained our earlier general conclusion in this special case that closure with respect to overriding external influence cannot be deduced from physical laws.)

This consideration can be carried over to deterministic laws, which can be treated as special case probabilistic laws (with a single possible outcome that is predicted to obtain with probability 1). If probabilistic laws are thought to attribute to physical systems intrinsic tendencies to produce outcomes, the same can be told of deterministic laws. But then, just as it was found impossible for an extrinsic influence to enter into conflict with the Born rule, it is equally impossible for it to contradict a deterministic physical law. So whoever is inclined to the propensity interpretation of probability, should also be inclined to see physical laws as ceteris paribus laws: they tell us what intrinsic tendencies there are in physical system to produce physical events, but they are silent about what should happen if a nonphysical influence obtains.

I think this way of thinking about physical laws should be natural for those who adhere to a tensed concept of time, involving that temporal determinations have an ontological import. On such a view of time, there is objective becoming, and what has already been laid out has powers to bring about what is not there yet. Somewhat loosely speaking, on such a picture God, as it were, would have to create only some initial state of the universe, plus the physical laws, and then any later stage of the universe would be brought about (either deterministically or probabilistically) by an earlier state. The laws of physics would report on the intrinsic tendencies in physical systems being in a certain state to produce later states. The reason that the physical properties that characterize different stages of the universe are lawfully connected would be that earlier stages had the power to bring about the later stages. On such a physical worldview, it is natural to think of physical laws as ceteris paribus laws. The supposition that maybe some external influence is capable of changing the course of physical events is completely irrelevant to the question what the intrinsic tendencies were in a previous state of the physical universe to bring about later physical events, therefore the falsity of such a supposition cannot be deduced from the laws of physics, whose only job is to report on such tendencies. (As Howard Stein (1968) has demonstrated, such a tensed view of time is in no conflict with the special theory of relativity – the language used in this paragraph would need to be adapted to the requirements of the equivalence of inertial observers, but objective becoming itself is admissible under relativity theory.)

On a tenseless view on time, on the other hand, past, present and future events (relative to one's frame of reference) are on a par, ontologically speaking. It makes no sense to attribute later events to the powers inherent in earlier physical states of affairs to bring them about. God, as it were, would have to create the whole four-dimensional manifold. Physical laws would be nothing over and above the fact that certain classes of events in the manifold display regularities. God would not be separately required to create physical laws, once the physical events are there. 'Causation' would be a word without much purchase in such a word, it would be just another, misleading, name for the fact that some collections of events subsume laws. A propensity interpretation of probability would be out of the question; if there should be objective probability, it would have to be grasped by some version of frequentism.

C) Is the brain deterministic at the neural level?

Some say the whole discussion in Subsection A is largely irrelevant because the brain is the only likely locus of interaction between the irreducibly mental and the physical, and even though indeterminism might prevail at the ground physical level, determinism prevails at the neural level. Ted Honderich, for one, claims that there is overwhelming evidence that the brain is a deterministic neural automaton:

> The first things to consider is neurons […]. Our mental lives are bound up with these most important elements of our brains and central nervous systems. Each of them is a cell into which go roots or dendrites […]. Out of each goes a trunk or axon […]. The roots are for

> input to the main body of the cell, and the trunk is for output. At the end of the trunk is a synapse or connection with other items, usually roots of other neurons. The input and output are electrochemical in nature. To begin with input to a root, chemical substances called neurotransmitters are released or secreted across a synapse, and this contributes to whether the neuron gets active or not. Some chemical inputs promote activity and some inhibit it. The activity is electrical and well understood. It consists in the passage of electrical impulses to the trunk of the neuron. These impulses occur in patterns, and result at the end of the trunk in the release of neurotransmitters across synapses to other neurons. A general truth about these building blocks of the brain and the nervous system is that their operation is indubitably taken to be causal [deterministic] by just about all working neuroscientists. No question can arise about that. (Honderich 2002: 65-6)

If it was indeed so, it would mean that as far as the causation of anything via the neuronal output of the brain is concerned, fundamental level quantum indeterminacies are completely irrelevant.

However, other philosophers, with equal assuredness in their tone, claim that virtually all working neuroscientists agree that quantum indeterminacies have an important role in the evolution of neural states, and therefore the brain is not a deterministic automaton. Henry Stapp, for one, writes that

> Quantum mechanics deals with the observed behaviour of macroscopic systems whenever those behaviours depend sensitively upon the activities of atomic-level entities. Brains are such systems. Their behaviours depend strongly upon the effects of, for example, the ions that flow into nervous terminals [synapses]. Computations show that the quantum uncertainties in the ion-induced release of neurotransmitter molecules at the nerve terminals are large (Stapp 1993: 133).
>
> These uncertainties propagate in principle up to the macroscopic level. Thus quantum theory must be used in principle in the treatment of the physical behaviour of the brain, in spite of its size (Stapp 2007: 300).

It is a somewhat unusual situation to have prominent and scientifically well-informed philosophers making contradictory claims, with equal levels of confidence, about an empirical matter. Since I have no expertise in the area, at this point I cannot presently do better than refer the reader to Mark Balaguer's work who, in the last few years, seemed to be surveying the state of play in neuroscience in this respect (Balaguer 2009, 2010, 2014, 2015). His conclusion, for the time being, is that it is "a wide open question," with little prospect to ever be settled by neuroscience (rather than physics), whether there are macro-relevant indeterminacies in neural processes.

In the latest and briefest statement of his findings, Balaguer cites a recent neuroscience textbook (Dayan and Abbott 2001), stating the following.

> [Synaptic] transmitter release is a stochastic process. Release of transmitter at a presynaptic terminal does not necessarily occur every time an action potential arrives and, conversely, spontaneous release can occur even in the absence of the depolarization due to an action potential. (p. 179)
>
> and
>
> Because the sequence of action potentials generated by a given stimulus varies from trial to trial, neuronal responses are typically treated statistically or probabilistically. For example, they may be characterized by firing rates, rather than as specific spike sequences. (p. 9)

He then goes on considering that some aspects of these (apparent) indeterminacies can be traced back to (apparent) indeterminacies in the opening or closing of ion channels, and argues that although it is not presently clear if these processes are genuinely indeterministic or merely appear so, it is clear that this question will not be answered by neuroscience, since it is presently clear that there is no deterministic neuroscientific explanation for them.

Time to draw conclusions for the whole section. I believe it has been demonstrated that the popular quick dismissal of the relevance of quantum indeterminacy for physical closure, on the ground that psycho-physical interaction in the gaps left in the physical causal explanation if some quantum processes are genuinely indeterministic "would disrupt the Born rule," are misguided. The most up-to-date surveys of neuroscience I am aware of seem to indicate that there is no neuroscientific evidence that the brain is a deterministic neural automaton, the question whether the ground level physical description of the particle-level processes that take place in the brain is deterministic or not may well be decisive in this question. Whether it is deterministic or not, as we have seen in Subsection A, is an open scientific question. In short, at present, there is no valid argument from determinism to closure with respect to law-conforming external influence, because there is no convincing empirical case for determinism (either at the basic physical, or the neural level).

8 Is there an inductive empirical (I-type) argument against overriding external influence?

The only question left to clarify is the one in the title of this section. The question of the title can be restated as this: Can we claim that (a) we looked for evidence for overriding external influence in the course of physical events, (b) we looked for it at all places where it is reasonable to expect that we would find it if overriding external influence really took place, (c) we employed methodologies that would be adequate to detect such evidence if there was any, and (d) there wasn't any, so (e) by normal inductive standards we are entitled to think that further investigation is not likely to discover any either?

Sometimes the retreat and "the fall" of British Emergentism is alluded to as a ground for an inductive argument for physical closure of exactly this sort. The British Emergentists posited special forces, of the sort discussed by Papineau (2001, see earlier), capable of 'downward causation' that pertain to the levels of reality that are the subject of the special sciences, such as chemistry, biology and psychology. However, the rise of quantum mechanics did away with special chemical forces, molecular biology reduced biological forces to chemical forces, which, in turn, are reduced to quantum mechanics, and, as it is claimed, neurophysiological research did the same to the alleged special mental forces. As it is emphasized in Brian McLaughlin's classic paper on British Emergentism (McLaughlin 1992), there was nothing incoherent about how Emergentists thought about the world, but the progress of science mounted overwhelming empirical evidence against their posits, the special forces. They posited breaches of physical closure at various places, we looked there carefully, and it turned out that there are none. We have every right, some might want to claim, in line with the normal standards of empirical science, to take it as direct inductive evidence for physical closure.

Maybe a word of caution is in order at this place. There is a pattern of bad inductive reasoning that should be avoided. It would go something like this. "I looked for my wallet in the kitchen, and it is not there. Same for the bedroom and the bathroom. By induction, I am satisfied that my wallet is not in the flat, although there is still the living room, where I never looked." We should make sure that we are not adopting the thesis of physical closure as a conclusion of an inductive argument that is no better than this one.

The most reasonable place to look for evidence for overriding external influence is the brain. Papineau claims "that neurophysiological research mapped the body's neuronal network and analysed the electrical mechanisms responsible for neuronal activity," and that "A great deal became known about biochemical and neurophysiological processes, especially at the level of the cell, and none of it gave any evidence for the existence of special forces not found elsewhere in nature."

Does it suffice? Recall, that Papineau's talk of "special forces" here relates to his argument, discussed earlier, to the effect that the conservation of energy would only allow for conservative law-governed external forces to affect the course of physical events. Even if we grant that the sort of empirical efforts Papineau refers to should have already identified such forces if there were any, it doesn't cut much ice, since, as we have seen earlier in Section 5, the argument to the effect that the conservation of energy rules out spontaneous external influence rests on a premise we have no reason to accept, since it has absolutely no support from physics, so we have no reason to believe the argument's conclusion. Consequently, what the proponent of the closure thesis would need is an inductive empirical argument to establish that there is no sign of external influence, influence that may well be spontaneous, not a subject to laws of any kind, in the course of physical brain events (in the succession of the physical states of the brain). Any physical event in the brain which is different from what is predicted by physical laws, taken together with previous physical events and external physical conditions, unless it can be attributed to inevitable noise in the experimental setup, could be considered as evidence for such interference. It is a huge empirical claim that there are no such events in the brain. It seems an impossibly hard experimental task to prove that at the ground physical level the physical state of the brain always evolves in a law-conforming way.

If there is objective physical indeterminacy involved in the course of physical events in the brain, and this indeterminacy propagates to the level of the neural determination of manifest action, which may well be the case as far as we can presently tell, what we would essentially need to get an empirical warrant against overriding externa influence is to make sure that in the quantum-mechanical processes that take place in the brain there is no diversion from the Born rule ever. To do that, we would have to observe large numbers of such quantum mechanical processes in which the relevant physical system (involving a part, or maybe the whole, of the brain) is prepared identically, and verify that the statistics of the outcomes observed conforms the Born rule. I think we have good reasons to doubt that it ever will be feasible, and I am sure that no such thing has ever been attempted.

9 Conclusions

I hereby rest my case that there is no scientific case for physical closure.

Of course, physical closure may be true. But it may also be false, for all we presently know. For example, if we are convinced, on the ground of reasons independent of physical closure, that physicalism holds, then we have every reason to believe in physical closure. The parsimony of physicalism is appealing, and if we think parsimony is a good guide of theory choice, it is reasonable to stick to physicalism unless we are forced by contrary evidence to abandon it.

But there seem to forceful considerations against physicalism. First and foremost, the place in an entirely physical world of the qualitative character of experience is entirely unclear. Finding a physicalist account for intentionality also seems highly problematic. The same can be said, I believe, of the freedom of the will and of moral responsibility, but of course we can live with notions of freedom and responsibility which are deeply revisionary of our respective pre-philosophical conceptions. Perhaps, as Epicurus once argued, if freedom and moral responsibility are on the line, so is the notion of intellectual responsibility—and whereas some philosophers might be willing to make peace with the thought that moral responsibility, as we naively conceived of it, is an illusion, letting go, or radically revising the meaning, of intellectual responsibility in a similar fashion would be the end of philosophy itself.

In this paper, I had no intention to argue for these points. My intention was to argue against the description of the situation, exemplified famously by Frank Jackson's abandonment of dualism, that, on the one hand, we have philosophical intuitions apparently favouring dualism, and on the other, we have overwhelming scientific evidence for physicalism (or for the disjunction of physicalism and epiphenomenalism, combined with strong philosophical considerations against the latter).

It is worth to take a short pause for Jackson' case.

Tellingly, he gave the title 'Epiphenomenal Qualia' to the paper in which he laid out the most discussed rendering of the knowledge argument (Jackson 1982), even though – as pointed out by Howard Robinson (2016) – the paper only presents arguments for the existence and the non-physicality of qualia, and none that would aim to establish that they are epiphenomenal. Instead, the last section of the paper contains responses to anticipated objections to qualia epiphenomenalism, as if it was obvious that if qualia exist and they are non-physical, then they are epiphenomenal, so any objection against their epiphenomenality was also an objection against the knowledge argument's conclusion.

Later, having long turned his back on the knowledge argument, in his "Mind and illusion" Jackson refers to the debate between physicalism and dualism as a clash between "what science tells us about the mind and its relation to the world" and "certain strongly held intuitions" that "suggest that there is something seriously incomplete about any purely physical story about the mind" (Jackson 2004: 421). In his "Postscript on Qualia," he considers that the responses he gave in the last section of "Epiphenomenal Qualia" to anticipated objections against qualia epiphenomenalism might not work against some of those objections. He comes to rejecting epiphenomenalism, after all, on the ground that

> Our knowledge of the sensory side of psychology has a causal source. Seeing red and feeling pain impact on us, leaving a memory trace which sustains our knowledge of what it is like to see red and feel pain on the many occasions where we are neither seeing red nor feeling pain. (Jackson 2004: 418.)

Then he goes on, within the same paragraph, declaring that

> We know […] that Mary's transition from not knowing what it is like to see red to knowing what it is like to see red will have a causal explanation in purely physical terms. (Dualist interactionism is false.)

This is a confession of faith in the closure thesis – the thesis which accounted for the presumptuous discussion of the defensibility of epiphenomenalism at the end of "Epiphenomenal qualia" in the first place. The picture that emerges from these texts is that the immutable fixed point in Jackson's thinking is the closure thesis, and it forced him to discard the knowledge argument's conclusion once he changed his mind about the respectability of epiphenomenalism.[6]

If there was a scientific case for closure, I agree with Jackson, responsible philosophers would be compelled to stick to science, and, in relation to arguments like the knowledge argument, the worthy task of philosophy would be to find out what has misled our intuitions, and how best we should defuse the seemingly forceful arguments for dualism, which we know must be unsound. The upshot of this paper is that there is no such scientific case at all.

Of course, the significance of this finding goes well beyond the question of the admissibility of interactive dualism. The whole point of the industry discussing the defensibility of nonreductive physicalism in the face of the causal exclusion argument (Kim 1989, 1998, 2005), goes out of the window, as well, along with a considerable part of the motivation of theorists who, feeling compelled by the arguments for the irreducibility of some mental phenomena they take to be non-epiphenomenal, but also believing that the emergentist option is blocked by the exclusion argument, adopt Russellian monist positions. As it has been noted before, the range of options in the philosophy of action is strongly affected, too.

Very briefly, this is how the argument of this paper went. First we recorded that there are two kinds of conceivable external influence in the course of physical events: one kind brings about physical

[6] For a formal reconstruction of the reasons Jackson gives for his apostasy in "Postscript…" and "Mind and illusion" see Robinson 2016: 57 ff.

events that appear to be in conformity with physical laws and earlier conditions, even though they may be different from what would have obtained in their place in the absence of external influence, the other brings about physical events that are ostensibly anomalous. We also distinguished between two kinds of possible arguments for physical closure: one kind would aim to deduce closure from the laws of physics (from particular physical laws, or form some general feature of some physical laws), the other would be an inductive argument on the ground that we made efforts to find physical events that cannot be accounted for the normal physical way, and none was found. As a second step, I argued that the two kinds of arguments and the two conceivable kinds of external influence are coordinated: deductive arguments have no teeth against overriding external influence, whereas inductive arguments on the ground of a lack of ostensibly anomalous physical events cut no ice against law-conforming external influence. Then we set out to examine if the deductive arguments proposed for closure (with respect to law-conforming external influence) are any good. We found that the most commonly used arguments from conservation principles suffer from quite elementary mistakes. Conservation principles would establish closure with respect to law-conforming external influence only if they entailed determinism, and they don't, except for very simple physical systems. Then we turned to the dynamical laws of physics to see if we have reason to believe in determinism, and found that physical determinism is an open scientific question which depends essentially on how quantum mechanics should be interpreted, or perhaps replaced by an even better theory. Arguments to the effect that quantum indeterminacy, supposing that an indeterministic interpretation prevails, would be irrelevant to the question of closure, because probabilistic dynamical laws would entail closure just as much as deterministic ones would do, were dismissed, mostly on the ground that they suffer from a failure to consider the meaning of physical probability. The suggestion that determinism might prevail at the neural level, nonetheless, was found to be an open scientific matter, too. Lastly, the claim that there is a strong enough inductive empirical argument against overriding external influence was discarded on the basis that there is no empirical case for the non-existence of the spontaneous kind of external influence, even if the empirical case for the non-existence of "special forces", i.e. the law-governed kind of overriding external influence, is convincing. The argument to the effect that spontaneous external influence could be excluded on the ground of the conservation of energy was found unsound.

We noted that a proponent of physical closure would be required to present arguments against both law-conforming and overriding external influence in the course of physical events to establish closure, so even if there was a sound deductive, or strong enough inductive, argument against one kind, the other kind should still be considered a possibility. But, as we have seen, there is no good argument against either kind.

The best prospect to improve on the, in its present state non-existent, case for physical closure would be if the physical causal order turned out to be deterministic, after all. At the present state of play it is certainly a possibility. Deterministic approaches to quantum mechanics, such as the various versions of Everettianism or Bohmian mechanics, appeal to many scholars. We briefly discussed the chances of obtaining empirical evidence favouring one or another of the competing interpretations, and recorded that, apart from dynamical collapse theories which differ from standard quantum mechanics in some of their testable predictions, so might in principle be differentially confirmed or disconfirmed, the rest of the competing interpretations agree in their predictions in the non-relativistic case, so they stand or fall together empirically, and even their relativistic generalizability is not a decisive guide to making a choice between them. But even if we had solid scientific reasons to favour a deterministic approach to quantum mechanics, thus to think that the physical causal order is deterministic, after all, there would still be the completely reasonable possibility that the laws of physics are best understood as ceteris paribus laws, and overriding external influence is possible. The experimental techniques that would be required to empirically exclude this possibility seem rather fantastic, and clearly beyond our present reach.

Returning, as a last note, to the theme of dualist interactionism, surely, none of the arguments presented in this paper brings us any closer to answering Princess Elisabeth's age-old question. I did not offer a positive account of how the non-physical mind is supposed to find a grip on physical reality to bring about a change in it. The lack of such an account, however, is not in itself an argument for physical closure, let alone a good one. We could be equally puzzled, for example, by the question how matter finds a grip on spacetime to make it curved. We know that matter does curve spacetime, and we know that the curvature of spacetime affects the distribution of matter, and we have the equations that account for the quantitative features of this dynamical interaction between two. It doesn't mean that we have an answer to the analogue of the Bohemian Princess' question in this particular case. Arguably, the limits of our understanding have been pushed much further in the matter-spacetime case than they have been in the mind-matter case, but there is, and probably there will always be, a residual mystery in the matter-spacetime case, too. We have obtained the laws that govern the interaction between them, and thereby spacetime is transformed from being the mere arena in which physical events take place to being part of the subject matter of physics. If there is an irreducible element of spontaneity in how a mind affects matter, then the mind will never be part, in its entirety, of any future extension of physics, and then it is futile to expect something like this in the case of the interaction between mind and matter. And then the mysteries are largely on a par.

References

Albert, David, and Barry Loewer (1988). "Interpreting the Many Worlds Interpretation," *Synthese* 77: 195-213.

Balaguer Mark (2009). "Why There Are No Good Arguments for Any Interesting Version of Determinism," *Synthese*, 168: 1-21.

Balaguer, Mark (2010). *Free Will as an Open Scientific Problem.* MIT Press.

Balaguer, Mark (2014). *Free Will.* MIT Press.

Balaguer, Mark (2015). "Determinism and its Relevance to the Free Will Question," in Jens Clausen and Neil Levy (eds.) *Handbook of Neuroethics.* Springer.

Bishop, Robert (2006). "The Hidden Premise in the Causal Argument for Physicalism," *Analysis* 66 (289):44-52.

Bohm, David (1952). *Quantum Theory*. Prentice-Hall Inc.

Bohr, Niels (1961). *Atomic Theory and the Description of Nature I: Four Essays with an Introductory Survey.* Cambridge University Press.

BonJour, Laurence (2010). "Against Materialism," in Robert C. Koons and George Bealer (eds.), *The Waning of Materialism.* Oxford University Press.

Broad, C. D. (1925). *The Mind and its Place in Nature*. London: Kegan Paul, Trench, Trubner & Co.

Chalmers, David (1996). *The Conscious Mind: In Search of a Fundamental Theory.* Oxford University Press.

Churchland, Paul M. (2014). "Consciousness and the Introspection of 'Qualitative Simples,'" in Rischard Brown (ed.) *Consciousness Inside and Out: Phenomenology, Neuroscience and the Nature of Qualitative Experience.* Springer.

Crane, Tim and D. H. Mellor (1990). "There is No Question of Physicalism," *Mind* 99 (394):185-206.

Dayan, Peter and Laurence F. Abbot (2001). *Theoretical Neuroscience: Computational and Mathematical Modeling of Neural Systems.* MIT Press.

de Broglie, Louis V. (1928) in *Electrons et Photons: Rapports et Discussions du Cinquième Conseil de Physique tenu à Bruxelles du 24 au 29 Octobre 1927 sous les Auspices de l'Institut International de Physique Solvay*, Paris: Gauthier-Villars.

DeWitt, Bryce S. (1971). "Quantum Mechanics and Reality: Could the solution to the dilemma of indeterminism be a universe in which all possible outcomes of an experiment actually occur?" *Physics Today*, 23(9): 30–40.

Dennett, Daniel C. (1991). *Consciousness Explained*. Boston: Little, Brown, and Company.

Deutsch, David (1996). *The Fabric of Reality*, New York: The Penguin Press.

Donald, Matthew J. (1990). “Quantum Theory and the Brain,” *Proceedings of the Royal Society* (London) Series A, Volume 427:43-93.

Donald, Matthew J. (1995). “A Mathematical Characterization of the Physical Structure of Observers”, *Foundations of Physics*, Volume 25: 529-71.

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

Eccles, John (1990). “A Unitary Hypothesis of Mind-Brain Interaction in the Cerebral Cortex,” *Proceedings of the Royal Society of London* B240: 433-51.

Einstein, Albert (1951). *Albert Einstein: Philosopher-Scientist,* edited by Paul Arthur Schilpp. The Library of Living Philosophers Inc.

Everett, Hugh (1957). “‘Relative State’ Formulation of Quantum Mechanics,” *Reviews of Modern Physics,* Volume 29, No.3: 454-62*.*

Gell-Mann, Murray and James B. Hartle (1990). “Quantum Mechanics in the Light of Quantum Cosmology”, in W. H. Zurek (ed.), *Complexity, Entropy and the Physics of Information.* Redwood City: Addison-Wesley, pp. 425–58.

Ghirardi, GianCarlo (2009). “The Interpretation of Quantum Mechanics: Where Do We Stand?” *Journal of Physics: Conference Series.* Volume 174, No.1.

Ghirardi, GianCarlo (2016). "Collapse Theories", *The Stanford Encyclopedia of Philosophy* (Spring 2016 Edition), Edward N. Zalta (ed.), URL = <http://plato.stanford.edu/archives/spr2016/entries/qm-collapse/>.

Ghiradi, GianCarlo, Philip Pearl, Alberto Rimini (1990). “Markov Processes in Hilbert Space and Continuous Spontaneous Localization of Systems of Identical Particles,” *Physical Review*, A42: 78.

Ghirardi, GianCarlo, Alberto Rimini, Tullio Weber (1986). “Unified Dynamics for Microscopic and Macroscopic Systems,” *Physical Review,* D34: 470.

Gisin, Nicolas (1989). “Stochastic Quantum Dynamics and Relativity,” *Helvetica Physica Acta*, 62: 363.

Graham, Neill (1973). “The Measurement of Relative Frequencies,” in DeWitt and Graham (eds.), *The Many Worlds Interpretation of Quantum Mechanics.* Princeton University Press.

Griffiths, Robert B. (1984). “Consistent Histories and the Interpretation of Quantum Mechanics,” *Journal of Statistical Physics*, 36: 219–72.

Griffiths, Robert B. (2013). “A Consistent Quantum Ontology,” *Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics* 44(2): 93-114.

Hájek, Alan (2009). “Fifteen Arguments Against Hypothetical Frequentism,” *Erkenntnis*, 70: 211–35.

Hendry, John (1984). *The Creation of Quantum Mechanics and the Bohr-Pauli Dialogue.* Reidel.

Honderich, Ted (2002). *How Free Are You?* Oxford: Oxford University Press

Huxley, Thomas (1874). “On the Hypothesis that Animals are Automata, and its History,” *Fortnightly Review* 95:555-80.

Jackson, Frank (1982). “Epiphenomenal Qualia,” *The Philosophical Quarterly* 32 (127): 127-36.

Jackson, Frank (2003). “Mind and Illusion,” in Anthony O’Hear (ed.), *Royal Institute of Philosophy Supplement,* Cambridge University Press.

Jackson, Frank (2004). “Postscript on Qualia,” in Peter Ludlow, Yujin Nagasawa and Daniel Stoljar (eds.), *There’s Something about Mary: Essays on Phenomenal Consciousness and Frank Jackson’s Knowledge Argument*. MIT Press.

Kim, Jaegwon (1989). “Mechanism, Purpose, and Explanatory Exclusion”, *Philosophical Perspectives*, 3: 77–108.

Kim, Jaegwon (1993). “Postscripts on Mental Causation”, in Kim, *Supervenience and Mind*. Cambridge University Press.

Kim, Jaegwon (1998). *Mind in a Physical World*. MIT Press.

Kim, Jaegwon (2005). *Physicalism, or Something Near Enough*. Princeton University Press.

Landau, Lev D., E. M. Lifshitz (1976). *Mechanics. Course of Theoretical Physics, Volume 1.* Third Edition. Elsevier.

Leibniz, Gottfried Wilhelm (1691/1896). “Essay on Dynamics on the laws of motion, in which it is demonstrated that not the same quantity of motion is preserved, but the same absolute force, or rather the same quantity of moving action, c. 1691,” Gottfried Wilhelm Leibniz, *New Essays Concerning Human Understanding*, translated by Alfred Gideon Langley. Macmillan.

Leibniz, Gottfried Wilhelm (1695/1896). “Essay on Dynamics in defence of the wonderful laws of nature in respect to the forces of bodies, disclosing their mutual actions and referring them to their causes, c. 1695,” Gottfried Wilhelm Leibniz, *New Essays Concerning Human Understanding*, translated by Alfred Gideon Langley. Macmillan.

Leibniz, Gottfried Wilhelm (1705/1989). “Considerations on Vital Principles and Plastic Natures, by the Author of the System of Pre-Established Harmony, 1705,” in Gottfried Wilhelm Leibniz, *Philosophical Papers and Letters,* edited by Leroy E. Loemker. Kluwer.

Lockwood, Michael (1996). “’Many Minds’ Interpretations of Quantum Mechanics,” *The British Journal for the Philosophy of Science,* 47:159-88.

Lowe, E. J. (2000). “Causal Closure Principles and Emergentism,” *Philosophy: The Journal of the Royal Institute of Philosophy*, 75 (294): 571–85.

McLaughlin, Brian P. (1992). “The Rise and Fall of British Emergentism,” in Ansgar Beckermann, Hans Flohr and Jaegwon Kim (eds.), *Emergence or Reduction? Prospects for Nonreductive Physicalism.* De Gruyter.

Omnés, Roland (1994). *The Interpretation of Quantum Mechanics.* Princeton University Press.

Omnés, Roland (1999). *Understanding Quantum Mechanics.* Princeton University Press.

Papineau, David (2001). “The Rise of Physicalism,” in Carl Gillett and Barry Loewer (eds.), *Physicalism and its Discontents.* Cambridge University Press.

Papineau, David (2002). *Thinking about Consciousness.* Oxford University Press.

Papineau, David (2009). “The Causal Closure of the Physical and Naturalism,” in Ansgar Backerman, Brian P. McLaughlin and Sven Walter (eds.), *The Oxford Handbook of Philosophy of Mind.* Oxford University Press.

Pearl, Philip (1976). “Reduction of the State Vector by a Nonlinear Schrödinger Equation”, *Physical Review*, D13: 857.

Pearl, Philip (1979). “Toward Explaining Why Events Occur”, *International Journal of Theoretical Physics*, 18: 489.

Popper, Karl (1957). “The Propensity Interpretation of the Calculus of Probability and the Quantum Theory,” in S. Körner (ed.), *The Colston Papers*, 9: 65–70.

Popper, Karl and John Eccles (1977). *The Self and Its Brain: An Argument for Interactionism.* Berlin, Heidelberg, New York: Springer.

Reichenbach, Bruce (1976). “Natural Evils and Natural Laws,” *International Philosophical Quarterly* 16 (2):179-196.

Robinson, Howard M. (2016). *From the Knowledge Argument to Mental Substance: Resurrecting the Mind*. Cambridge University Press.

Saunders, Simon (1993). “Decoherence, Relative States, and Evolutionary Adaptation,” *Foundations of Physics*, 23: 1553-85.

Saunders, Simon (1995). “Time, Quantum Mechanics, and Decoherence,” *Synthese*, 102: 235–66.

Stapp, Henry (2007). "Quantum Mechanical Theories of Consciousness", in Velmans, M. and S. Schneider (eds.), *The Blackwell Companion to Consciousness*. Oxford: Blackwell.

Stapp, Henry (1993). *Mind, Matter and Quantum Mechanics*, New York: Springer.

Stein, Howard (1968). "On Einstein-Minkowski Space-Time." *Journal of Philosophy*, 65: 5-23.

Tegmark, Max (1997). "The Interpretation of Quantum Mechanics: Many Worlds or Many Words?" in M. H. Rubin and Y. H. Shih (eds.), *Fundamental Problems in Quantum Theory*. Proceedings of UMBC Workshop.

von Mises, Richard (1957). *Probability, Statistics and Truth*. New York: Macmillan.

von Neumann, John (1955). *Mathematical Foundations of Quantum Mechanics*. Princeton: Princeton University Press.

Wallace, David (2012). 2012, *The Emergent Multiverse*. Oxford University Press.

Woolhouse, R. S. (1985). "Leibniz's Reaction to Cartesian Interaction," *Proceeding of the Aristotelian Society* 86:69-82.

Wigner, Eugene P. (1997) The Collected Works of Eugene Paul Wigner: Part A. The Scientific Papers III, Part I. Particles and Fields; Part II. Foundations of Quantum Mechanics. Ed. By A. S. Wightman. Berlin.

Wigner, Eugene P. (1967) *Symmetries and Reflections.* Cambridge, MA.: MIT Press.