If Consciousness Causes Collapse, the Zombie Argument Fails

Mousa Mohammadian (mmohamma@nd.edu)

This is a pre-print of an article published in *Synthese*. The final authenticated version is available online at: https://doi.org/10.1007/s11229-020-02828-4.

Abstract Many non-physicalists, including Chalmers, hold that the zombie argument succeeds in rejecting the physicalist view of consciousness. Some non-physicalists, including, again, Chalmers, hold that quantum collapse interactionism (QCI), i.e., the idea that non-physical consciousness causes collapse of the wave function in phenomena such as quantum measurement, is a viable interactionist solution for the problem of the relationship between the physical world and the non-physical consciousness. In this paper, I argue that if QCI is true, the zombie argument fails. In particular, I show that if QCI is true, a zombie world physically identical to our world is impossible because there is at least one law of nature, a fundamental law of physics in particular, that exist only in the zombie world but not in our world. This shows that philosophers like Chalmers are committing an error in endorsing the zombie argument and QCI at the same time.

Keywords the zombie argument, zombies, laws of nature, interactionism, collapse

1. Introduction

Physicalists hold that consciousness is either physical or can be reductively explained in terms of physical entities and processes—e.g., in terms of brain and neural processes. The zombie argument is one of the most well-known anti-physicalist arguments. It relies on the conceivability of zombies: creatures that are physically identical to conscious beings but lack consciousness. According to the zombie argument, we can conceive a world that is physically identical to our world but instead of conscious beings, it is filled with their zombie twins. If such a world is conceivable, it is metaphysically possible and if it is

metaphysically possible, consciousness is not physical. If it was, a world physically identical to our world without consciousness wouldn't be possible. Now, if consciousness is non-physical, what is its relationship with the physical world? Historically, one major approach to address this question is interactionism. It is the idea that consciousness and the physical world are causally efficacious on one another: conscious states cause physical states and physical states cause conscious states.

There is a debate in the literature about the consistency of the zombie argument and interactionism. John Perry (2001, pp. 72–77, 2012) famously argues that if interactionism is true, some physical events are caused by consciousness in our world. Since there is no consciousness in the zombie world, either these physical events do not exist or, if they exist, they are physically caused. Anyway, the zombie world would not be physically identical to our world and hence the zombie argument fails. In response, Chalmers (2004) suggests that the zombie world has causal gaps. That is, the physical events that are caused by consciousness in our world do happen in the zombie world but without cause. This proposal aims to keep the two worlds physically identical without damaging the conceivability of the zombie world.

In this paper, I argue that Chalmers' proposal does not succeed. For this reason, I focus on (arguably) the most prominent version of interactionism, recently advocated by Chalmers and McQueen (forthcoming), that I call *Quantum Collapse Interactionism* (QCI). Briefly, it is the idea that consciousness causes the collapse of the wave function in phenomena such as quantum measurement. I argue that if QCI is true, the zombie world is physically different from our world and therefore the zombie argument fails.

The structure of the paper is as follows. Section 2 briefly discusses the zombie argument. In Section 3, I explain QCI in detail. In this paper, I do not discuss the independent plausibility of the zombie argument or QCI. Rather, in Section 4, I show that if QCI is true, the zombie argument fails because the zombie world and our world would be physically different. In particular, I show that there is a law of nature in the zombie world that does not exist in our world. Finally, in Section 5, I consider two objections against my argument and show that they do not succeed in eliminating the inconsistency between the zombie argument and QCI.

2. The Zombie Argument

The zombie argument[1] relies on the conceivability of zombies. My zombie twin is my exact physical replica, molecule-by-molecule identical to me, who shares all my functions and behavioral dispositions, but *lacks consciousness*. When I touch a very sharp object, for instance, my nervous system reacts, I move my hand quickly, and say 'ouch' loudly. My zombie twin demonstrates all these reactions too. However, whereas I feel pain, my zombie twin does not. According to the proponents of the zombie argument, we can imagine a zombie world which is physically identical to our world—i.e., all the physical facts about it are the same as all the physical facts about our world—but everyone in it is the zombie twin of their twin conscious being in our world. The zombie argument states that if the zombie world (or my zombie twin) is conceivable, then it is metaphysically possible. But this means that consciousness should be a non-physical component of our world. "If God could have created a zombie world, then [...] after creating the physical processes in our world, he had to do more work to ensure that it contained consciousness" (Chalmers 2003, p. 106). Therefore, the argument goes, consciousness is non-physical.

3. Quantum Collapse Interactionism (QCI)

Some believe that quantum mechanics provides us with new resources to think about the relationship between the physical and the mental. In particular, some physicists and philosophers suggest that a phenomenon in quantum measurement called *the collapse of the wave-function* is evidence for interactionism.[2] To clarify this proposal, let's begin with an ordinary case of measurement or observation.

Imagine that there is a very large bag full of small balls. The bag is made of thick fabric so we cannot see the color of the balls when they are in the bag. For millions of times, whenever we brought a ball out of the bag, its color was either black or white. Under normal circumstances, we commonly concede that the color of a ball that is observed to be black (or white) was already black (or white) when it was in the bag, namely, before observing its color. But imagine we realize that when the balls are still in

[1] Different versions of the zombie argument can be found in the literature—for an extensive list, see (Kirk 2015). Here, I focus on Chalmers (1996, pp. 94–99, 2003, pp. 105–106, 2010).

[2] There are several proposals about the relationship between consciousness and quantum physics. Two interesting summaries reflecting the diversity of these proposals can be found in Atmanspacher (2017) and Pylkkänen (2018). It should be mentioned that some of these proposals are physicalist. Penrose (1989, 1994), for instance, aims to explain consciousness in terms of quantum physics in general and collapse of the wave-function in particular.

the bag, they are not simply either black or white. Rather, they are either black, or white, or some shade of grey such as silver grey, elephant grey, magnetic grey, asphalt grey, and so on. But once we bring them out of the bag and observe their color, they always turn into either black or white. This is a very simple illustration of what happens in quantum measurement. When, for instance, we measure or observe the spin of an electron, it is always either up or down. However, our best physical theory indicates that in many situations, *before conducting the measurement*, the spin is not merely either up or down. Rather, it might be in a state called a superposition of states that is not up or down but something "in-between," similar to the shades of grey that are "in-between" black and white. The process of shifting from a superposition of states to a single measured state (called an eigenstate), which happens at the moment of measurement, is called the collapse of the wave-function (henceforth simply "collapse").

The standard theory of quantum mechanics tells us that the superposition of states collapses into eigenstate at the moment of measurement. Yet, it does not tell us what exactly constitutes measurement (this is the well-known quantum measurement problem). Different interpretations of quantum mechanics propose different answers to the quantum measurement problem. The Copenhagen Interpretation, considered by many as the orthodox interpretation of quantum mechanics, suggests that quantum measurement can be understood in terms of the interaction between microscopic objects (e.g., electrons) that obey quantum mechanical laws and macroscopic measuring devices that are "entirely described in irreducibly classical terms" (Schlosshauer 2007, p. 27). In this view, the borderline between the quantum world and the ordinary world—which is called *the Heisenberg cut*—separates the domain in which an electron's spin can be in a superposition of states from the domain in which it can only be in the classic-like eigenstate of "either up or down."

In his very influential *Mathematical Foundations of Quantum Mechanics* (2018/1932), von Neumann claims that although we have to accept the Heisenberg cut and "divide the world into two parts, the one being the observed system, the other the observer" (2018/1932, 272), there is no unarbitrary way to decide where exactly the cut and hence the moment of measurement (and collapse) should be placed. We can place it between microscopic objects and measuring devices, or between measuring devices and the human observer, or on the observer's retina, or optic nerve, or brain. According to von Neumann, in this latter case, the observer is the physicist's "abstract "ego"" (2018/1932, 273) and the observed system is her brain, optic fiber, retina, measuring devices, and

microscopic objects altogether. Based on von Neumann's account of quantum measurement, some physicists have suggested that we could push the Heisenberg cut all the way up towards the observer such that everything made out of atoms (and hence material) becomes a part of the observed system. Here, the only thing that is left to be the observer is non-material consciousness that causes the collapse (London and Bauer 1983/1939; Stapp 2001, 2007; Wigner 1961, 1964). This is the gist of QCI according to which non-physical conscious intervention causes collapse.[3]

In *The Conscious Mind* (1996), Chalmers considers, but does not advocate, QCI as a candidate for explaining the relationship between the physical world and the non-physical consciousness. In a later work, he claims that there is no "knockdown argument" against QCI (1999, pp. 492–3). Recently, his view of QCI has become even more enthusiastic. He (2003, pp. 125–6) claims that quantum mechanics is "perfectly compatible" with QCI and "positively encouraging" its possibility. Moreover, similar to the proponents of QCI, he holds that "it is natural to suggest that a measurement is precisely a conscious observation, and that this conscious observation causes a collapse." Finally, in their "Consciousness and the Collapse of the Wave Function" (forthcoming), Chalmers and McQueen examine QCI in detail and endorse it as a viable research program to explain the relationship between the physical and the mental.

Here, we do not need to go through the details of their model of QCI and a general picture is sufficient for the purpose of our current discussion. In their view, conscious states *nomologically* supervene on quasi-classical brain states. That is, there are some fundamental "physics-to-consciousness" law(s) stating how consciousness arises from physical processes of some parts of the brain called "physical correlates of consciousness" (PCC). When PCC enters into a superposition of states, since consciousness is always in alignment with PCC, it enters into the superposition too. Yet, consciousness is superposition-resistant and if it gets superposed, it rapidly collapses. The collapse of consciousness, according to a "consciousness-to-physics" law, results in the collapse of PCC which, in turn, results in the collapse of the superposed system that PCC is a part of. For instance, when an experimenter measures the spin of a superposed electron, as a result of the interaction between the

[3] It should be mentioned that not all interpretations of quantum mechanics are "collapse interpretations." For instance, Everett's many-worlds interpretation and Bohm's hidden variables interpretation do not embrace collapse. Moreover, QCI is only one of the collapse interpretations—and not a very popular one (Schlosshauer et al. 2013). For instance, Ghirardi-Rimini-Weber (GRW) theory of spontaneous collapse is also a collapse interpretation but consciousness plays no role in it.

electron and the measuring device, the device gets "entangled" with the electron and enters into the same superposition of states. Similarly, as a result of the interaction between "the electron and the measuring device" system and the experimenter's retina, the retina enters into the superposition of states and so on until the PCC gets superposed. However, PCC's superposition results in the superposition of consciousness. At this point, consciousness rapidly collapses and causes the collapse of PCC (and every physical object that is entangled with it) into its eigenstate.

In the following, I argue that if QCI is true, the zombie argument fails because there can't be a zombie world physically identical to our world. Thus, between the zombie argument and QCI, one can only select one.

4. Can there Be a Zombie World Physically Identical to Our World?

Let's suppose that QCI is true in our world. That is, let's suppose that consciousness is non-physical and it causes collapse. To run the zombie argument, three zombie worlds can be taken into account:

(1) A zombie world in which collapse occurs with a physical cause;

(2) A zombie world in which collapse does not occur;

(3) A zombie world in which collapse occurs uncaused.

As mentioned before, John Perry (2001, pp. 72–77, 2012) provides general arguments showing that if interactionism is true, worlds similar to (1) and (2)—that is, zombie worlds in which the physical effects of consciousness either do not exist, or they are caused physically—are not physically identical to our world. Therefore, they cannot be used to run the zombie argument. In the case of collapse, in (1), collapse has a physical cause. Therefore, there is a physical causal relationship in (1) that does not exist in our world. If we adopt Frank Jackson's definition of physical facts, that is, "everything in *completed* physics, chemistry, and neurophysiology, and all there is to know about the causal and relational facts consequent upon all this" (1986, p. 291), then there is a physical fact, corresponding to this physical causal relation in (1) that does not exist in our world. In (2), collapse never occurs. So the result of an experiment in which a zombie physicist measures the spins of some electrons is always different from the result of an identical experiment done by her twin human physicist. This, again, constitutes a physical difference between (2) and our world. So the only remaining option, which is endorsed by Chalmers (2004), is (3): a zombie world in which collapse occurs exactly as it does in our world but without any cause.

In the following, I argue that if QCI is true, there is a physical fact about (3) that is not true in our world. Therefore, even if the zombie world is like (3), it is still physically distinguishable from our world. (Henceforth, by 'the zombie world' I always mean (3).)

4.1. A Zombie World with Uncaused Collapses

As mentioned before, assuming that QCI is true in our world, according to Chalmers and McQueen's account, the following fact about collapse in the human world is true:

FC_H Always collapse occurs at the moment that consciousness enters (or is about to enter) into a superposition of states.

If FC_H is the most basic fact about collapse in the human world, which seems a reasonable claim, in the zombie world, where there is no consciousness, the following statement can be considered its corresponding fact about collapse:

FC_Z Always collapse occurs at the moment that PCC[4] enters into a superposition of states.

Prima facie, in the zombie world, FC_Z looks like the perfect equivalent of FC_H in the human world. First, it is a purely physical fact that does not cite any non-physical or psychophysical entity or relation. Second, it is a common fact between the zombie world and the human world because it is true in both worlds. Finally, FC_Z does not add any entity or causal relation that does not exist in the human world to the zombie world. The only major difference between FC_H and FC_Z is that the former cites the cause of collapse, namely, consciousness. FC_Z, however, is merely a "brute fact" about collapse in the zombie world that does not draw a causal relation between PCC and collapse because in the zombie world collapse is *uncaused*. Overall, however, if FC_Z is true, collapse occurs in the zombie world as it occurs in the human world and hence the two worlds are physically identical.

I do not think that this proposal succeeds in eliminating all physical differences between the zombie world and the human world. In the following, I argue that FC_Z, if true, is a law of nature—a fundamental law of physics, in particular—only in the zombie world but not in our world. This means that there is at least one physical fact about the zombie world that is not a fact of our world. Therefore, these two worlds are not physically identical.

[4] The physical correlates of consciousness (PCC) in a zombie is the part of her brain that if she was her human twin, her consciousness would nomologically supervene on it.

4.2. FC_H and FC_Z: Three Similarities

Assuming that FC_H and FC_Z are true in their worlds, some interesting and interrelated similarities emerge between them. First, they both can be used to *explain* instances of collapse in their worlds. Consider, for example, a particular instance of collapse that occurs when an electron's spin is measured. In the human world, when the system that includes the superposed electron entangles with PCC and, as a result, the experimenter's consciousness is about to get superposed, collapse happens. Here, FC_H provides a proper answer for "Why did the system collapse?" Facing a similar question in the zombie world, one can properly say "Because always collapse occurs at the moment that PCC enters into a superposition of states."

Prima facie, one might think that only FC_H is a proper explanation for particular instances of collapse because it cites consciousness which is the cause of collapse. But FC_Z does not cite the cause of collapse. Quite conversely, it implies that it is uncaused and hence it does not really explain particular instances of collapse in the zombie world. There are two problems with this claim. First, citing a cause for an event is not the only way to explain it. Most major recent accounts of scientific explanation are open to the possibility of non-causal explanations for particular events, partially because there are many cases of non-causal explanations in science (Bokulich 2018; Lange 2017; Woodward 2005, p. 221, 2018). Granted, FC_Z does not cite any cause for an instance of collapse in the zombie world. Yet, it can be considered a non-causal explanation (e.g., a structural explanation) for collapse, in the same way that "Nothing can travel faster than the speed of light" explains an electron's subluminal velocity without citing any cause for it.[5]

Secondly, one might even argue that FC_Z is indeed a causal explanation for instances of collapse in the zombie world. According to David Lewis (1986, p. 217), "*to explain an event is to provide some information about its causal history*" (original italics). In his view, saying that an event is uncaused still provides some information, albeit *negative* information, about the event's causal history by stating that it is empty. If one admits this view, FC_Z can be considered a causal explanation for instances of collapse in the zombie world exactly because it implies that collapse is uncaused. Therefore, FC_Z can be used as an

[5] Woodward (2005, pp. 208–09), for instance, argues that "Nothing can travel faster than the speed of light" is not a causal explanation for the subluminal speed of an electron (cf. Skow 2014, pp. 455–57).

explanation (whether non-causal or causal) for instances of collapse in the zombie world. This constitutes the first interesting similarity between FC_H and FC_Z.

The second interesting similarity between FC_H and FC_Z is that they both support relevant counterfactuals in their corresponding worlds. Imagine, for instance, that a human physicist and her zombie twin are about to measure the spin of the superposed electron *e*. However, their brain does not interact with the superposed system of the physical entities that are entangled with *e* because, for instance, they leave the lab just before the interaction happens. Would the superposed system collapse if it had entangled with their brain? In the human world and in the zombie world, we can respond positively to this question in virtue of FC_H and FC_Z, respectively.

It is quite obvious that FC_H supports the counterfactual claim. If the system had entangled with the human physicist's brain, his consciousness would enter into a superposition of states. This, however, would result in the collapse of the system. Yet, when it comes to the zombie world, one might object to the claim that FC_Z supports the counterfactual "If the system had entangled with the zombie physicist's brain (and hence his PCC), it would collapse." First, it might be said that the counterfactual is true but *not* in virtue of FC_Z. Rather, there is another proposition, true in the zombie world, which supports this counterfactual. My answer to this objection is simple but, I believe, strong. First, the burden of proof is on the objector to tell us what true proposition in the zombie world supports the counterfactual. Second, it is very unlikely for the objector to find any such proposition. After all, in the zombie world, the only relevant fact that we know about collapse and its relationship with the brain in general and PCC in particular is FC_Z.

Secondly, the objector might deny the truth of the counterfactual altogether. Namely, she might say that in the zombie world, it is not true that if the superposed system had been entangled with the zombie physicist's brain, it would collapse. If the counterfactual is false and FC_Z is true, then FC_Z cannot support the counterfactual. This claim is problematic too. If we adopt the possible worlds approach in interpreting counterfactual conditionals, the counterfactual is true if and only if in all of the most similar possible worlds to the zombie world in which the system entangles with PCC, it collapses. Thus, the counterfactual is false if there is a possible world, most similar to the zombie world, in which if the system entangles with PCC, it does not collapse. Such a possible world should "depart" from the zombie world after the system's entering into a superposition but before zombie physicist's leaving the lab. So

she stays in the lab, the system entangles with her brain, and hence her PCC enters into a superposition of states, but it does not collapse. Since this possible world is perfectly similar to the zombie world up to after the system's entering into a superposition, it has the same laws of nature as the zombie world. Yet, collapse does not occur when PCC enters into the superposition. This means that "superposed PCC without collapse" is physically possible in a world with the same laws of nature as the zombie world. Therefore, it should be physically possible in the zombie world too.

Obviously, if such a possibility gets actualized, that is, if a zombie physicist's brain entangles with a superposed system but doesn't collapse, we have a physical difference between the zombie world and our world. To avoid this problem, we should admit that although "superposed PCC without collapse" is physically possible in the zombie world, it never happens. In this regard, "superposed PCC without collapse" is just like many other physically possible things that never happen. It is physically possible for me to throw my laptop out of the window or to speak Esperanto. But, as much as I am concerned, they never happen. I never throw my laptop out of the window because it will be destroyed. I never speak Esperanto because I have no reason to learn it and one cannot speak a language without learning it. Similarly, the objector might say that "superposed PCC without collapse" is physically possible in the zombie world but it just never happens. There is, however, a problem here. To say that an event that never happens is physically possible is to say that there is some physically attainable condition under which the event can actually happen. I never throw my laptop out of the window but there is some physically attainable condition under which I will do that (e.g., if an aggressor is out there who wants to severely hurt me and throwing my laptop at him is my only choice to save my life). I never speak Esperanto but there is some physically attainable condition under which I will learn and speak Esperanto (e.g., if I have a daughter who loves Esperanto and I realize that speaking Esperanto with her is a great way to strengthen our relationship). Similarly, if "superposed PCC without collapse" is physically possible in the zombie world, there must be some physically attainable condition under which a zombie's PCC can be superposed without collapsing. This, however, means that if under such condition we compare the brain of the zombie physicist with that of her human twin, there is a physical difference between them. To avoid this problematic consequence, we should admit that the counterfactual claim is true in the zombie world. That is, FC_Z supports relevant counterfactuals in the zombie world.

The third interesting similarity between FC_H and FC_Z pertains to how they receive confirmation from their instances. When a human physicist hypothesizes that FC_H is true, how can she convince other physicists that she is correct? She probably needs to provide some theoretical grounds for her claim, similar to what Chalmers and McQueen do in their work, and conduct some experiments to show that collapse really occurs as a result of conscious intervention. She might succeed or not but, anyway, she is not expected to test *all* possible cases of collapse to convince her colleagues that FC_H is true. In this regard, FC_H is unlike the claim that "All the coins in my pocket are nickels." This claim can be known to be true only if I check *all* the coins in my pocket and see that they are nickels.

The situation is similar for FC_Z. Imagine, for instance, that a zombie physicist proposes that FC_Z might be true and a group of zombie experimental physicists conduct some experiments and observe that collapse really occurs when PCC is superposed. Do zombie physicists go ahead and conduct quantum experiments that rely, in one way or another, on the assumption that collapse will occur according to FC_Z? For instance, do they go ahead and examine whether large systems can also enter into a superposition and collapse? Do they conduct experiments to detect collapse of the wave function under different circumstances? Or, rather, every time they hesitate to conduct such experiments because they are worried that despite all previous cases, maybe collapse ceases to occur when PCC enters into a superposition? If zombie physicists' behavior and scientific practice resemble that of their human twins[6] (and they should), they would simply assume that collapse will occur at that moment. That is, they would take previous cases of collapse when PCC is superposed as *confirming evidence* for future untested cases of collapse at this moment without thinking that they need to test *all* possible cases of superposed PCC to see whether FC_Z is true. Therefore, both FC_H and FC_Z are taken to be well-established before an exhaustive enumeration of all their instances. Thus, they can be used to predict future cases of collapse too.

[6] As noted by many, whether or not a claim receives confirmation from evidence is always sensitive to the background *beliefs* (see, for instance, Lange 2000, p. 112; Sober 1988, p. 19). If we adopt representationalism, zombies cannot have beliefs because beliefs are some kind of mental state. Therefore, they cannot really confirm their beliefs. But since, *ex hypothesi*, zombie's behavior perfectly resembles their human counterparts, here I adopt the dispositionalist view of belief according to which beliefs should be understood in terms of behavioral dispositions. For more on this distinction, see Schwitzgebel (2019).

4.3. FC_Z and the Characteristics of Laws of Nature

If my arguments are sound, the universal regularity FC_Z has three important characteristics in the zombie world: (1) it explains particular instances of the regularity, (2) it supports corresponding counterfactuals, and (3) it receives instantial confirmation from its observed instances and hence can be used for making predictions about unobserved instances.

But there is a *nearly unanimous consensus* in the literature on laws of nature that these characteristics demarcate laws of nature from accidental regularities. Humeans (Lewis 1973; Loewer 1996, p. 11), non-Humeans (Armstrong 1983, Chapter 4; Bird 1998, Chapter 1; Dretske 1977; Tooley 1977, 1987, p. 57), antireductionists (Lange 2000, pp. 11–23), and even antirealist (van Fraassen 1989, Chapter 2) *all* agree that only laws of nature—but not accidental regularities—instantiate these characteristics. One (or probably *the*) major reason for this consensus is that these characteristics are essential in how laws are actually used in scientific practice. Disagreements among the proponents of different accounts of laws of nature are primarily about explaining *how* and *why* laws of nature instantiate such nomic characteristics while, for instance, accidental generalizations fail to do so—or, in the case of antirealists, whether there is anything that really instantiates such characteristics. Now, since FC_Z satisfies all the above-mentioned conditions, according to all major accounts of laws of nature, *FC_Z is a law of nature* (a law of physics, in particular).

To argue otherwise, that is, to argue that FC_Z is not a law of nature but a mere accidental regularity, one needs to show that as opposed to what I argued for, either FC_Z does not satisfy the above-mentioned conditions or, more strongly, that satisfying those conditions does not make it a law of nature. The latter simply amounts to a whole new view of laws of nature, dramatically different from our current philosophical and scientific understanding of laws and their properties. The former, I believe, seems like a daunting task too, especially because we have to model the zombie world after our own world. Given QCI, since FC_H is true in the human world, we have to assume that FC_Z is true in the zombie world. Moreover, even in the human world, FC_Z is not only true but also the physical part of the law-governed psychophysical mechanism of collapse. Consequently, as I have shown, we have very strong intuitions that FC_Z manifests uniquely nomic characteristics such as explaining instances of collapse, supporting relevant counterfactuals, and receiving confirmation from its instances.

But FC_Z is not just a law of physics. Rather, it is a *fundamental* law of physics. It instantiates the characteristics that Chalmers mentions for fundamental laws, namely, marking "the end of the explanatory chain" (1996, p. 76), describing fundamental features of the world (1996, p. 126), simplicity (1996, p. 127), and being clear and precise (1996, pp. 338–39). FC_Z is simple, it cannot be explained and has to be taken as a brute fact about the zombie world. It describes collapse, which is a fundamental physical process that, together with Schrödinger's equation, constitute the "core of quantum mechanics" (Chalmers 1996, p. 336). FC_Z is also clear and precise. In fact, if we finally succeed in constructing a complete and precise psychophysical theory of consciousness—as Chalmers suggests—then we shall know what PCC exactly is and, as a result, all the current ambiguities of FC_Z will be eliminated.

But FC_Z does not manifest the characteristics of laws of nature in the human world. First, it does not support relevant counterfactuals in the human world. For instance, given QCI, although counterfactual "If a superposed system had entangled with Sarah's brain, it would collapse" is true in our world, it is not true *in virtue of* FC_Z. Rather, it is true in virtue of the truth of the following proposition:

a) "If a superposed system had entangled with Sarah's brain, her PCC would be superposed;"
b) "If Sara's PCC had been superposed, her consciousness would be superposed;"
c) "If Sara's consciousness had been superposed, it would collapse;"
d) "If Sara's consciousness had been collapsed, her PCC would collapse."

Here, (a) is true because it is supported by a law of physics and the other counterfactuals are true because they are supported by psychophysical laws posited by Chalmers and McQueen's model of QCI. Therefore, "If a superposed system had entangled with Sarah's brain, it would collapse" is not true in virtue of FC_Z.

Second, in the human world, FC_Z does not explain instances of collapse. It is obviously not a causal explanation because it does not cite the cause of collapse, i.e., conscious intervention. Moreover, FC_Z does not provide a non-causal explanation for collapse either. A complete argument for this claim goes beyond the scope of this paper because to show that FC_Z is not a non-causal explanation for collapse, one needs to show that it does not fit the characteristics of any types of non-causal explanation or to show that collapse is not a kind of fact that admits such explanations. So, here, relying on the most comprehensive account of non-causal explanations in the literature (Lange 2017), I simply provide a brief overview.

Some non-causal explanations explain an event by identifying some particular constraint with which the world must comply. This constraint can be mathematical or physical. For instance,[7] a mother always fails to distribute exactly 23 strawberries evenly among her 3 children (without cutting strawberries) *because* 23 cannot be divided evenly into whole numbers by 3. Or, one might argue, an electron does not travel faster than the speed of light *because* nothing can travel faster than the speed of light (that is, the speed of light is a physical constraint for speed). Another type of non-causal explanation is what Lange calls a "Really Statistical" explanation which shows that the explanandum is "just a statistical fact of life" (2017, p. 189). An example of a "Really Statistical" explanation is explaining a phenomenon by regression toward the mean. Now, it is quite obvious that FC_Z does not identify a mathematical constraint to which the world must conform. It does not identify a physical constraint either. If it was, that is, if collapse's occurring at the moment of PCC's entering into a superposition was a physical constraint to which the world must comply, then collapse should have occurred at this moment *even without conscious intervention*. This, however, is inconsistent with QCI. Finally, the phenomenon of collapse is not just a statistical fact, nor is FC_Z a "Really Statistical" explanation. Thus, FC_Z is not a non-causal explanation for collapse in our world.

Therefore, FC_Z does not instantiate the essential characteristics of laws of nature in our world. As a result, it is only a law of nature in the zombie world but not in our world.

5. Two Objections

So far, I have argued that FC_Z is a fundamental physical law only in the zombie world. Thus, this world is not physically identical to ours. And if the two worlds are not physically identical, the zombie argument fails. In the following, I respond to two objections that aim to address the problem of inconsistency between the zombie argument and QCI in different ways.

5.1. Objection One

Any physicalist will agree that a conscious state at a given time *t* supervenes on some physical facts in a period that extends from some time before *t* up to *t*. For instance, according to physicalists, the state of my consciousness at the moment of writing these words supervenes on some physical facts within a period that extends from some time in the past up to the moment of writing these words. How long this

[7] The example is from Lange (2017, p. 6).

period is should not concern us here. Now, imagine a new zombie world that is physically identical to our world only until an arbitrary time in the future (t^*). After t^*, the new zombie world physically differs from our world such that collapse no longer occurs. According to the proposal, this new zombie world can be used to run the zombie argument because it is conceivable; it is physically identical to our world until t^*; and, as a result, it includes all the physical facts that a physicalist might hold that our current conscious states supervene upon. Moreover, in this new zombie world, FC_Z is false because after t^* collapse will not occur when PCC is superposed. Thus, FC_Z is not a law of nature in the new zombie world and hence it cannot constitute a physical difference between our world and the new zombie world. This means that we can accept the zombie argument and QCI without falling into inconsistency.

I do not think that this proposal poses a serious challenge for us. We just need to introduce a new version of FC_Z:

FC_Z^* Until t^*, always collapse occurs at the moment that PCC enters into a superposition of states.

FC_Z^* is *always* true in the new zombie world (even after t^*), it indicates a regularity in this world, it explains particular instances of this regularity, it supports corresponding counterfactuals, it admits instantial confirmation from its instances and can be used for prediction. Therefore, it is a law of nature in the new zombie world. However, it is not a law in our world and hence it constitutes a physical difference between the two worlds.

It might be suggested that FC_Z^* cannot be a law of nature, because a law cannot contain any ineliminable references to specific individuals, times, places, events, and so on. In FC_Z^*, however, there is an ineliminable reference to a particular time that limits its application to a specific period of the zombie world's history. Therefore, FC_Z^* is not a law of nature. There are two problems with this claim. First, as Lange (2000, pp. 34–39) argues, there is no reason to think that there is such a requirement for laws of nature. Regardless of the problems with cashing out what exactly this requirement amounts to, a law can play all of its scientific roles even if it includes a local predicate. FC_Z^* should be considered a law of nature because it has all the functions and essential characteristics that laws of nature have in science, even though it includes a local predicate. Secondly, and more importantly, the requirement that a law must cover the whole history of its corresponding world is especially ill-grounded in a world in which nature's behavior is subject to change. The new zombie world is such a world. In this world, there is no

"superposed PCC without collapse" up to t^* but afterward, collapse does not occur. If the physical behavior of a world changes at some point in its history, it shouldn't be a surprise that its laws of nature change accordingly.[8]

5.2. Objection Two

According to the second objection, FC_Z's being a law of nature in the zombie world is not really problematic for the zombie argument. The objection goes as follows:

> *Premise 1*. The zombie world and our world must be physically identical *only* vis-à-vis the physical facts on which consciousness might supervene.
>
> *Premise 2*. Physical facts about collapse are not among the facts on which consciousness might supervene.
>
> *Conclusion*. Therefore, the zombie world and our world don't have to be identical vis-à-vis the facts about collapse (e.g., FC_Z).

Consequently, FC_Z's being a law of nature in the zombie world but not in our world does not render the zombie world unsuitable for running the zombie argument against physicalism.

Here, the objector accepts that if we adopt QCI, the zombie world is not identical to our world but claims that this disanalogy is not problematic for the zombie argument because the two worlds are identical regarding the physical facts upon which consciousness might supervene. But this objection succeeds only if the objector can provide us with a sufficiently specific theory of consciousness according to which facts of collapse are not among the physical facts upon which consciousness supervenes. Let's see if Chalmers has such a theory at his disposal.

Chalmers and McQueen (forthcoming) structure their consciousness-collapse model around Tononi's (2008; 2016) integrated information theory of consciousness. However, as they explain, this specific theory is not essential for their model. Rather, their approach "can be generalized to any psychophysical theory [of consciousness] linking *quasi-classical* states [of the brain] to states of consciousness" (emphasis added). In this picture, one might argue that facts of collapse, which belong to the domain of microphysical, are not included among the physical facts upon which consciousness supervenes because "quasi-classical" PCC does not really belong to the domain of microphysical. This

[8] It is worth mentioning that some hold that even in the course of *our* world's history, laws of nature change and evolve (Peirce 1891; Smolin 2013).

claim, however, is not obviously correct. First, PCC is physical, can be superposed, and since it is related to consciousness, it collapses. Thus, in the set of all physical facts about PCC, there are some facts about collapse. Second, in Chalmers and McQueen's model, the so-called Heisenberg cut is located either on PCC or between PCC and consciousness. Anyway, PCC is still in the domain of the quantum rather than the classic. These do not necessarily mean that these "collapse facts" are among the physical facts on which consciousness supervenes. For consciousness might supervene on a subset of all physical facts about PCC which does not include any collapse fact. But it also does not mean that the collapse facts are definitely excluded from this subset. To make the difference between the zombie world and our world irrelevant for the zombie argument, more works should be done to show that collapse facts are certainly not among the physical facts on which consciousness supervenes.

But even if in Chalmers and McQueen's theory of consciousness (CMTC), consciousness does not supervene on physical facts about collapse, zombie argument is still in trouble. As Chalmers and McQueen acknowledge, CMTC can be interpreted in two ways. It can be viewed dualistically when consciousness is taken to be non-physical and only nomologically supervening on PCC. Or it can be viewed physicalistically when consciousness is, for instance, identical to PCC. Let's call the former "the dualist CMTC" and the latter "the physicalist CMTC." They argue that both versions of CMTC can be used to provide good (and nearly empirically identical) consciousness-collapse models that provide us with empirically testable dynamics of collapse. Although the physicalist CMTC is ontologically more parsimonious—it only posits physical properties—they prefer the dualist CMTC because "we already have good reason to believe that consciousness is a fundamental [non-physical] property." What is this "reason"? I assume that Chalmers' response includes the collection of anti-physicalist arguments that he has offered in his works.

Now, imagine that relying on the "original" zombie argument—which requires a zombie world physically identical to our world—we adopt the dualist CMTC and Chalmers and McQueen's model for QCI. However, as a result of adopting QCI, a discrepancy emerges between the zombie world and our world regarding the FC_Z. So we modify the zombie argument adding that it does not matter if the zombie world is not identical to our world regarding the facts of collapse because, according to CMTC, facts of collapse are not among the physical facts on which consciousness might supervene. Thus, a "revised" zombie argument according to which the two worlds need not be identical regarding the facts

about collapse, e.g., FC_Z. But recall that we have used the "original" zombie argument—which does not work anymore—to support selecting the dualist CMTC rather than its physicalist rival. So now we need to go back and see whether the "revised" zombie argument can also be used to support this choice. Unfortunately, it cannot. For the "revised" zombie argument only works if we *presuppose* that facts about collapse are not among the physical facts on which consciousness might supervene. But this presupposition, in turn, is accepted only because of CMTC. To sum, if we accept CMTC, we do make the "revised" zombie argument non-problematic but then we cannot rely on the zombie argument to select the dualist CMTC rather than its physicalist rival.

An obvious solution might be suggested for this problem: the zombie argument is only one of the many anti-physicalist arguments. We can invoke another argument, e.g., the knowledge argument, to support adopting the dualist CMTC instead of the physicalist CMTC. Once we adopt this theory, we go ahead and develop our QCI model and now we can show that the "revised" zombie argument works despite the fact that the two worlds are different regarding FC_Z. This approach eliminates the inconsistency between QCI and the "revised" zombie argument but at a great cost: we can accept the "revised" zombie argument only *after* adopting non-physicalism relying on other arguments such as the knowledge argument. In other words, to eliminate the inconsistency between the "revised" zombie argument and QCI, this approach makes the zombie argument an obsolete anti-physicalist argument.

6. Conclusion

I argued that if QCI is true, that is, if collapse of the wavefunction is caused by consciousness, then there is a physical fact about the zombie world that is not a physical fact about our world. In particular, there is a fundamental law of physics in the zombie world that does not exist in our world. This argument is not, at least directly, an argument for physicalism—although it makes it harder for non-physicalists to explain the relationship between the mental and the physical by removing one of their options. Rather, it shows that philosophers like Chalmers are committing an error in endorsing the zombie argument and QCI at the same time.

But the same conclusion can be seen from a more interesting perspective. Imagine that physicists and philosophers of physics follow Chalmers and McQueen's proposal and adopt QCI as a research program to find, among other things, an answer for the quantum measurement problem. Moreover, imagine that this research program turns out to be impressively successful such that most

physicists and philosophers of physics accept it as the best solution for the quantum measurement problem. This might look like good news for the proponents of the zombie argument because it confirms the non-physicality of consciousness. Yet, it is not. As I argued, if QCI is true, the zombie argument fails. This would be a quite ironic situation: if the most well-known non-physicalist solution for the quantum measurement problem succeeds, the most well-known argument for non-physicalism fails.

References

Armstrong, D. M. (1983). *What is a law of nature?* Cambridge: Cambridge University Press.

Atmanspacher, H. (2017). Quantum approaches to brain and mind. In S. Schneider & M. Velmans (Eds.), *The Blackwell companion to consciousness* (pp. 298–313). Hoboken, NJ: John Wiley & Sons, Ltd.

Bird, A. (1998). *Philosophy of science*. London: UCL Press.

Bokulich, A. (2018). Searching for non-causal explanations in a sea of causes. In A. Reutlinger & J. Saatsi (Eds.), *Explanation beyond causation: philosophical perspectives on non-causal explanations* (pp. 141–163). Oxford: Oxford University Press.

Chalmers, D. J. (1996). *The conscious mind: in search of a fundamental theory*. New York: Oxford University Press.

Chalmers, D. J. (1999). Materialism and the metaphysics of modality. *Philosophy and Phenomenological Research*, *59*(2), 473–496.

Chalmers, D. J. (2003). Consciousness and its place in nature. In S. P. Stich & T. A. Warfield (Eds.), *Blackwell guide to the philosophy of mind* (pp. 102–142). Malden: Blackwell.

Chalmers, D. J. (2004). Imagination, indexicality, and intensions. *Philosophy and Phenomenological Research*, *68*(1), 182–190.

Chalmers, D. J. (2010). The two-dimensional argument against materialism. In *The character of consciousness* (pp. 141–205). Oxford: Oxford University Press.

Chalmers, D. J., & McQueen, K. J. (forthcoming). Consciousness and the collapse of the wave functions. In S. Gao (Ed.), *Consciousness and quantum mechanics*. Oxford: Oxford University Press.

Dretske, F. I. (1977). Laws of nature. *Philosophy of Science*, *44*(2), 248–268.

Jackson, F. (1986). What Mary didn't know. *The Journal of Philosophy*, *83*(5), 291–295.

Kirk, R. (2015). Zombies. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Summer 2015.). Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/sum2015/entries/zombies/. Accessed 26 October 2017

Lange, M. (2000). *Natural laws in scientific practice*. Oxford: Oxford University Press.

Lange, M. (2017). *Because without cause: non-causal explanation in science and mathematics*. New York: Oxford University Press.

Lewis, D. (1973). *Counterfactuals*. Malden, MA: Wiley-Blackwell.

Lewis, D. (1986). Causal explanation. In *Philosophical papers* (Vol. 2, pp. 214–240). Oxford: Oxford University Press.

Loewer, B. (1996). Humean supervenience. *Philosophical Topics*, *24*(1), 101–127.

London, F., & Bauer, E. (1983). The theory of observation in quantum mechanics. In J. A. Wheeler & W. H. Zurek (Eds.), *Quantum theory and measurement* (pp. 217–259). Princeton: Princeton University Press.

Peirce, C. S. (1891). The architecture of theories. *The Monist*, *1*(2), 161–176.

Penrose, R. (1989). *The emperor's new mind: concerning computers, minds, and the laws of physics*. Oxford: Oxford University Press.

Penrose, R. (1994). *Shadows of the mind: a search for the missing science of consciousness*. Oxford: Oxford University Press.

Perry, J. (2001). *Knowledge, possibility, and consciousness*. Cambridge, MA: MIT Press.

Perry, J. (2012). Return of the zombies? In S. Gozzano & C. S. Hill (Eds.), *New perspectives on type identity: the mental and the physical* (pp. 251–263). Cambridge: Cambridge University Press.

Pylkkänen, P. (2018). Quantum theories of consciousness. In R. J. Gennaro & P. Pylkkänen (Eds.), *The Routledge handbook of consciousness* (pp. 216–231). New York: Routledge.

Schlosshauer, M. (2007). *Decoherence and the quantum-to-classical transition*. Berlin: Springer. Accessed 27 March 2019

Schlosshauer, M., Kofler, J., & Zeilinger, A. (2013). A snapshot of foundational attitudes toward quantum mechanics. *Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics*, *44*(3), 222–230.

Schwitzgebel, E. (2019). Belief. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Fall 2019.). Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/fall2019/entries/belief/. Accessed 3 December 2019

Skow, B. (2014). Are there non-causal explanations (of particular events)? *The British Journal for the Philosophy of Science*, *65*(3), 445–467.

Smolin, L. (2013). *Time reborn: from the crisis in physics to the future of the universe*. Boston: Houghton Mifflin Harcourt.

Sober, E. (1988). Confirmation and law-likeness. *The Philosophical Review*, *97*(1), 93–98.

Stapp, H. (2001). Quantum theory and the role of mind in nature. *Foundations of Physics*, *31*(10), 1465–1499.

Stapp, H. (2007). Quantum mechanical theories of consciousness. In S. Schneider & M. Velmans (Eds.), *The Blackwell companion to consciousness* (pp. 300–312). Hoboken, NJ: John Wiley & Sons, Ltd. Accessed 7 March 2019

Tononi, G. (2008). Consciousness as integrated information: a provisional manifesto. *The Biological Bulletin*, *215*(3), 216–242.

Tononi, G., Boly, M., Massimini, M., & Koch, C. (2016). Integrated information theory: from consciousness to its physical substrate. *Nature Reviews Neuroscience*, *17*(7), 450–461.

Tooley, M. (1977). The nature of laws. *Canadian Journal of Philosophy*, *7*(4), 667–698.

Tooley, M. (1987). *Causation: a realist approach*. Oxford: Oxford University Press.

van Fraassen, B. C. (1989). *Laws and symmetry*. Oxford: Oxford University Press.

von Neumann, J. (2018). *Mathematical foundations of quantum mechanics*. Princeton: Princeton University Press.

Wigner, E. P. (1961). Remarks on the mind-body question. In I. J. Good (Ed.), *The scientist speculates* (pp. 168–181). London: Basic Books.

Wigner, E. P. (1964). Two kinds of reality. *The Monist*, *48*(2), 248–264.

Woodward, J. (2005). *Making things happen: a theory of causal explanation*. New York: Oxford University Press.

Woodward, J. (2018). Some varieties of non-causal explanation. In A. Reutlinger & J. Saatsi (Eds.), *Explanation beyond causation: philosophical perspectives on non-causal explanations* (pp. 117–138). Oxford: Oxford University Press.