Mental Causation

Kobašlić, Matej

SVEUČILIŠTE U RIJECI

FILOZOFSKI FAKULTET U RIJECI

ODSJEK ZA FILOZOFIJU

Mental Causation

Can mental properties exert their causal powers in a physical world?

Diplomski rad

MENTOR: Dr. sc. Boran Berčić

STUDENT: Matej Kobašlić

Rijeka, srpanj 2019.

Table of Contents

SAŽETAK

Kroz većinu rada bavit ću se problemom mentalnog uzrokovanja, odnosno pitanjem mogu li mentalna svojstva imati kauzalne moći u fizičkom svijetu. Točnije, bavit ću se problemom kauzalne ekskluzije: budući da fizička svojstva *isključuju* mentalna svojstva u kauzalnim procesima, mentalna svojstva su kauzalno nepotrebna. Nakon analiziranja različitih mogućnosti rješavanja navedenog problema, cilj posljednjeg dijela rada jest ponuditi interpretaciju modela redukcije koju je iznio Jaegwon Kim. Prema tom modelu, mentalna svojstva mogu imati kauzalne moći ukoliko ih se identificira s njihovom funkcionalnim realizatorima. Ujedno smatram da je ovaj model trenutno najbolja opcija koju imamo u rješavanju problema mentalnog uzrokovanja.

U prvom poglavlju ovoga rada izložit ću motive za raspravu, odnosno pokazati zašto nam je mentalno uzrokovanje potrebno. Nakon toga slijedi kratki uvod u Descartesov dualizam supstancije i njegovim pokušajem rješavanja problema interakcije između uma i tijela. Dualizam supstancije ne uspijeva riješiti problem budući da krši princip *fizičke zatvorenosti*. Prikazat ću argument kauzalne ekskluzije i dilemu mentalnog uzrokovanja: ili mentalna svojstva imaju kauzalne moći, ili ih nemaju. Ukoliko ih nemaju, onda su *epifenomenalna*, odnosno kauzalno impotentna. Ukoliko ih imaju, onda postoji mogućnost da krše princip fizičke zatvorenosti. Prvo ćemo vidjeti zašto epifenomenalizam nije održiva opcija. Nakon toga analizirat ću opcije prema kojima možemo očuvati mentalno uzrokovanje. Prva grupa opcija će biti *redukcionističke*, točnije *teorija identiteta* i *materijalni eliminativizam.* Obje će te opcije biti odbačene. To nas dovodi do druge opcije: *ne-reduktivni fizikalizam*, a posebno Davidsonov *anomalni monizam*. Međutim, pokazat će se kako obje mogućnosti u konačnici dovode do problema kauzalne ekskluzije. Treća će opcija također biti ne-reduktivna te pokazuje na koji način *kontrafaktička analiza kauzalnosti* može očuvati mentalnom uzrokovanju. No, ispostavit će se da ni ta opcija neće biti u potpunosti uspješna. U posljednjem poglavlju prikazat ću Kimov model *funkcionalne redukcije* i nastojat ću pokazati kako nam taj model može pomoći u očuvanju mentalnog uzrokovanja.

Ključne riječi: mentalno uzrokovanje, kauzalna ekskluzija, epifenomenalizam, redukcionizam, anti-redukcionizam, funkcionalna redukcija, Jaegwon Kim

ABSTRACT

The aim of this thesis is twofold. For the main part, the aim is to deal with the problem of mental causation, answering the question of whether mental properties can have causal powers in a physical world. More precisely, it will deal with the problem of *causal exclusion*, according to which physical properties *exclude* mental properties in causal processes, making them causally inefficacious and inert. For the final part, after analysing different options that could resolve this issue, my aim is to provide an interpretation of Jaegwon Kim's model of reductionism. According to this model mental properties *do have* causal powers because they are identified with their functional realizers. I will argue that this might be the best option we currently have in saving mental causation.

In the first chapter I will explain motivation for mental causation, answering the question of why do we need mental causation at all? This will be followed by a rather brief preliminary introduction to Descartes' substance dualism and his struggle to resolve the problem of mind-body interaction. Having seen how dualism fails to resolve the problem due to the principle of *physical closure*, I will present the argument of causal exclusion and the dilemma of mental causation: either mental properties have causal powers, or they do not. If they do not, they are *epiphenomenal*, i.e. causally impotent. If they do, they might be in danger of breaking the principle of physical closure. First, we will see why epiphenomenalism is not a tenable option. Then we turn to options that save mental causation. The first group of options will be *reductionist* in nature, namely *identity theory* and *eliminative materialism*, both of which will be rejected. This leads us to the second option: *non-reductive physicalism*, Davidson's *anomalous monism* in particular. However, we will see that these two options in fact generate the problem of causal exclusion. The third option is also non-reductive, and it shows how *counterfactual analysis of causation* might save mental causation, but such an undertaking falls short from being successful. For the final chapter, Kim's model of *functional reductionism* will be presented and analysed. I will try to show that this model saves mental causation.

Keywords: mental causation, causal exclusion, epiphenomenalism, reduction, non-reduction, functional reductionism, Jaegwon Kim

Volitions do not enter into the chain of causation… The feeling that we call volition is not the cause of a voluntary act, but the symbol of that state of the brain which is the immediate cause.

– T. H. Huxley, 1874

1. INTRODUCTION

1.1. Explaining behaviour

What does it mean to say that my thoughts and my intentions caused me to behave in a specific manner? For example, what caused me to type down the previous sentence? Well, someone could easily argue that linguistic knowledge that I possess together with my wanting to express a thought that's been boggling me caused me to type it down. A simpler example would be the following: I want to raise my hand and I do so. It seems rather hard to doubt these processes, as we use similar examples in our everyday life. Almost everything that we do, we attribute it to our thoughts, wishes, even qualitative mental states; when we go to see a doctor, we usually explain it by saying that something is hurting us, or that we are in pain. Moreover, psychology in its core rests on explanations in terms of our mental states. This is the key component of *mental causation*, the idea that our mental states have causal powers.

However, the philosophical issue central to this thesis occurs when we start to interpret our behaviour and aforementioned causal processes using knowledge of neurosciences, medicine, physiology, physics, and other similar branches of science. Now, more than ever before, we have mapped out our brain processes to such detail that, at least in principle, every causal process can be explained in terms of the physical properties. We can explain our behaviour using neurosynaptic relations, prefrontal cortex, C-fibres, etc. What does this mean exactly? Let us take an example from Norman Malcolm (1968), whose work served as an inspiration to much of contemporary literature that I shall analyse in this thesis. Imagine a man climbing a ladder leading to the roof of his house. We also know that a wind blew off his hat which landed onto the same roof. What is the explanation for his behaviour, i.e. climbing the ladder to the roof? One explanation is rather simple, provided in terms of mental states as seen in the first paragraph. Malcolm spells it out in the following form:

> If a man wants to retrieve his hat and *believes* this requires him to climb a ladder, he will do so provided there are no countervailing factors.
> This man *wanted* to retrieve his hat and *believed* that this required him to climb a ladder, and there were no countervailing factors.
> Therefore, he climbed a ladder. (Malcolm 1968: 48; emphasis added)

The emphasis in this explanation is put on his beliefs and wanting to do something. Yet, we can also explain his behaviour by appealing to his neurophysiology, firing of neurons, chemical changes in body tissue, etc.[1] (p. 52)

According to Malcolm (p. 52-3), the exact problem occurs when we have a collision of two explanations, or accounts: mental and physical, especially because the physical, i.e. neurophysiological theory is supposed to provide *sufficient* causal explanation. If his behaviour is *completely* accounted for physically, what is left for mental, or intentional? Malcolm argues that:

> given the antecedent neurological states of his bodily system together with general laws correlating those states with the contractions of muscles and movements of limbs, he would have moved as he did regardless of his desire or intention. If every movement of his was completely accounted for by his antecedent neurophysiological states (his "programming"), then it was not true that those movements occurred *because* he wanted or intended to get his hat. (Malcolm 1968, p. 53)

Even more general upshot of the argument is that the *physical excludes the mental* when explaining certain behaviour.

And thus, we reach the main question of this thesis: how is it possible for the mental to exert causal powers in the world that is fundamentally physical?[2] If physical processes are causally sufficient for our behaviour, what purpose does the mental have? It is easy to realise why Malcolm's work was so influential, as it would be hard to accept that our behaviour is nothing more but a consequence of a bunch of neurophysiological processes happening in our brain and our body. As Heil and Mele (1993) in their preface state, "we confront a dilemma. Either we concede that 'purposive', reason-giving explanations of behaviour have only a pragmatic standing, or we abandon our conception of the physical domain as causally autonomous." (Heil and Mele 1993 p. v)

[1] Clearly, such an explanation requires more finesse and detail than provided here, but the fundamental idea stands.
[2] Paraphrased from Kim (1998).

1.2. What do we need mental causation for?

Before I continue with unravelling the issue of mental causation, one has to explain why many philosophers find this problem to be of utmost importance, as I share the same motivation with them. To get the full understanding, we need to answer the following question: what is it about mental causation that needs desperate saving? Why do we even need it? First and foremost, mental causation as such requires one thing: that mental properties causally *make difference*. If mental properties are causally superfluous, as Kim's argument will show, then the consequences might be dire. One way to go about our need for mental causation lies in human agency and moral responsibility.[3] The case is quite simple, as we have seen in the introduction. In order to explain our behaviour, we usually appeal to our mental states. Why am I writing this paper? Well, I want to graduate. In order to graduate, I know I need to write my thesis. These reasons are rather extrinsic, but I also have an intrinsic motivation; I am curious about the development of sciences, its implications to philosophy, and its future. My desires to know more, my intentions to graduate, my beliefs of what I need to do in order to achieve my goals – all of these elements seem crucial for my action, i.e. writing this thesis. Not only that, but if it weren't for mental causation, I wouldn't be able to write this thesis in a coherent and logically structured manner, since mental causation also has consequences on our memory and reasoning. As Kim argues: "if you take away perception, memory, and reasoning, you pretty much take away all of human knowledge." (1998: p. 31)

In other words, if it weren't for mental causation, it would be hard to recognise ourselves as moral agents and it would be hard to attribute any moral values to our actions. This might seem like a far-fetched idea, but there are philosophers who infer even more radical consequences. Jerry Fodor (1989) believes that:

> If it isn't literally true that my wanting is causally responsible for my reaching, and my itching is causally responsible for my scratching, and my believing is causally responsible for my saying . . . , if none of that is literally true, then practically everything I believe about anything is false and it's the end of the world. (p. 156)

Although I find this argument to be more of a hyperbole rather than a cataclysmic depiction of the end of the world, Fodor's reasoning still shows this intuition that philosophers (and even philosophical laymen alike) almost unanimously share. In what follows, we will see in precise

[3] I shall present similar arguments as Kim (1998), although further references can be found in Robb and Heil (2019).

detail what exactly threatens mental causation and how different philosophers have worked to resolve it. But first, let us briefly see how it all actually started.

1.3. Descartes, substance dualism, and Princess Elisabeth

There are two elementary things that need to be unpacked here. First, according to Descartes, what exists? To rephrase the question, what substances exist? Second, what is the relation between those elements? To simply answer the first one: there are two substances. As we know, Descartes famously starts his *Meditations* ([1641] 1993) with a "doubt", calling into question everything he knows, presuming he might be tricked. The aim of this doubt is to see whether there is something that cannot be questioned, one thing that is certainly and necessarily true. For Descartes, that certainty lies in his *cogito, ergo sum*, "thought exists; it alone cannot be separated from me. I am; I exist – this is certain" (*Second Meditation*, p. 19/27). This leads him to the conclusion that there exists a thinking thing, a thinking substance, *res cogitans*. But there is also another substance, although one that can be brought into doubt. In the same chapter Descartes posits the existence of another substance, the material, corporeal and extended one, *res extensa*, out of which the physical world is constructed.[4]

That answers the first question. Now, what about the relation between the mind and body, i.e. the material? The answer to this question might be unclear. First, in *Sixth Meditation*, he claims that the thinking thing and the physical, extended thing are radically different; "(…) it is certain that I [the thinking thing] am really distinct from my body, and can exist without it." (p. 51/78). According to Descartes, the distinctness stems from the following properties: the thinking thing, i.e. the mind, is immaterial, lacking spatial dimensions, "utterly indivisible", meaning it does not have parts, and it is indestructible by natural means – in other words, it is eternal. Body, on the other hand, is completely opposite – material, has spatial dimension, divisible and destructible. Yet, he claims that there is a "union", the "commingling of the mind with the body" (p. 53/81), and even in the *Sixth Meditation* Descartes provides examples of causal interaction concerning sensation of pain (belonging to the mind) and its physical manifestation (p. 57/87).

[4] To provide this brief and rather simplistic version of Descartes' substance dualism, I have helped myself with Hatfield's article on René Descartes from *Stanford Encyclopedia of Philosophy.*

However, if that is the case, how exactly does the mental causally influence the physical? It seems rather unclear and even implausible that such a causal relation can even occur. Following the same intuitions, in 1643, Princess Elisabeth of Bohemia posed a problem to René Descartes' substance dualism in a way that even Descartes himself struggled to find an adequate answer. In a letter to Descartes, Princess Elisabeth states the following:

> How the mind of a human being can determine the bodily spirits [i.e., the fluids in the nerves, muscles, etc.] in producing voluntary actions, being only a thinking substance. For it appears that all determination of movement is produced by the pushing of the thing being moved, by the manner in which it is pushed by that which moves it, or else by the qualification and figure of the surface of the latter. Contact is required for the first two conditions, and extension for the third. [But] you entirely exclude the latter from the notion you have of the body, and the former seems incompatible with an immaterial thing.[5]

If these two substances are so radically different and opposite, with no contact whatsoever, it seems rather unclear how one could causally influence the other. To sum it up more formally:

P1: The mind and the body are two entirely different substances.

P2: Entirely different substances cannot interact with one another[6].

C: Mind and body could not interact.[7]

1.4. Completeness of the Physical or Physical closure

What does this tell us? Is there anything to be learned from Descartes? One might argue that Cartesian dualism is a thing of the past, something that not many philosophers believe to be a matter of fact, thus rendering this discussion irrelevant. I believe there is something to be learned from Descartes. It seems to me that one valuable lesson, or principle that we can extrapolate is *Physical Closure*, a principle that Cartesian dualism tends to break. Philosophers

[5] Elisabeth to Descartes, May 1643. This quotation is taken from Garber (2001: p. 172).

[6] This premise seems to be of utmost interest; Descartes thought this was not the case, but it is questionable whether he could argue otherwise successfully. One of the options he opted for later on in his work was introduction of the pineal gland, stating: "when the mind wants to remember something, this volition makes the gland incline successively in different directions" (cf. Garber 2001: p. 145-6). But even this interpretation seems to be unable to resolve the exact nature of the causal mechanisms, and as I am going to argue later in this paper, similar objections can be raised to non-reductive physicalism.

[7] This argument is also endorsed by Sarah Patterson (2005), who provides a more historical overview of the discussion.

almost unanimously agree that Descartes is in fact breaking a *principle of Energy Conservation* (e.g. Dennett 1991; Fodor 1981). For example, Heil (2013) states that:

> Mass and energy are convertible, but the total amount of mass–energy is constant. Mind–body interaction threatens to violate conservation. Intervention from the 'outside' by a mind would increase total energy without a compensating loss of mass, and mass–energy would be lost were material bodies to affect minds. (p: 26)

Later in this paper we will see that the principle of physical closure still plays a very important role in contemporary philosophy. The consequence of the principle is the following, as states by Kim:

> If you pick any physical event and trace out its causal ancestry or posterity, that will never take you outside the physical domain. That is, no causal chain will ever cross the boundary between the physical and the nonphysical. (Kim 1998: p. 40)

If this principle is something that we could all agree on, and Kim together with many other philosophers (e.g. Robb and Heil 2019; Heil 2013) seem to argue that we really should accept it, then we face a dilemma that might not be so easy to solve.[8]

1.5. Causal Exclusion

Within Cartesian picture, the problem of mental causation occurs when substance dualism meets the scientific criteria of energy conservation, or physical closure. But the problem does not end there. We have already seen how Malcolm stresses out the problem of having two *explanations* and how one (physical) excludes the other (mental). Since explanations typically trace causation, the problem holds for *causal powers* as well. For that reason, what I am going to present now is the (in)famous *causal exclusion* argument, most famously developed by Jaegwon Kim (1998; 2005). The details of the argument, together with its motivation and Kim's solutions are going to be presented later. For the purposes of this chapter, I will just provide a definition accompanied with an example. Kim defines the principle of causal exclusion in the following fashion:

[8] It needs to be noted that there were interpretations of Descartes that explained the relation of mind and body in a dualist sense. His contemporary, Malebranche, believed in occasionalism. Just briefly, occasionalism posited that the interaction occurred by appealing to the action of God (Patterson 2005: p. 248). However, even these interpretations seemingly render mental causation to be irrelevant, i.e. not having any true causal powers.

> *Principle of causal exclusion*. If an event *e* has a sufficient cause *c* at *t*, no event at *t* distinct from *c* can be a cause of *e* (unless this is a genuine case of causal overdetermination). (Kim 2005: 17)

Before moving on, I need to briefly explain what *overdetermination* is. The concept of *overdetermination* implies cases where there are two *sufficient* causes. For example, two bullets hitting victim's heart at the same time (Kim 2005: p. 48), or two kids throwing rocks at a window simultaneously. These cases rarely happen, but they are nonetheless possible. The idea is that both causes can bring about the effect, independent of each other.

What does this tell us about mental causation? For a start, let us imagine a situation from the beginning of the paper: me wanting to raise my hand and me doing so successfully. One interpretation of this case would be to say that my *wanting* to raise my hand *caused* me to do so. However, if we stick to Cartesian sense of "my wanting" as something non-physical, we break the principle of physical closure. This principle also implies that my raising of the hand is caused by something physical and neurosciences can explain it in terms of some neurosynaptic processes from my prefrontal cortex, through my body, all the way to the muscles in my arm, etc. If this is the case, *do my mental states play any role in this process?* If neurons or some other neurophysiological processes and states do all the causal work, what is left for my mental states? They seem to be *causally inefficacious* and *inert*, and this is central claim of *epiphenomenalism* – a position according to which we have mental states that are *caused* by some physical states, but they in turn do not have any causal powers. As Stoljar defines it, it is "a version of dualism according to which mental events and properties have no causal role in the production of physical events." (Stoljar 2010: p. 234)

This "conclusion" is something that we could hardly accept. So, what options do we have? This generates a dilemma (Bregant, 2003), horns or which shall be the basis for this work:

> Either mental properties have causal powers od they do not. If they have them, we risk a violation of the causal closure of the physical domain; if they do not have them, we embrace epiphenomenalism, which denies any sort of causal powers to the mental. So, either we violate the causal closure of physics, or we end up with epiphenomenalism. (Bregant 2003: 305)

These two horns of the dilemma are going to be tackled in the following sense. First, I will reject epiphenomenalism as an option, leaving the first horn, which claims that mental properties *do* have causal powers. This horn will in fact have two more options; a *reductive*

and *non-reductive physicalism*; either we adopt a reductionist theory which *saves* mental causation, but at a cost of *reducing* mental properties to physical or we can opt for non-reductive physicalism and provide a theory that could *save* mental causation *without having to reduce* mental properties to physical. But first, as I said, I will try to avoid the threat of epiphenomenalism that the causal exclusion argument seems to posit.

1.6. The Threat of Epiphenomenalism

First and foremost, what needs to be pointed out is that although this argument seemingly leads to epiphenomenalism, most of the authors (namely Kim 1998, or Yablo 1992) consider this argument to be *reductio ad absurdum.* Epiphenomenalism is option that we tend to reject, leaving us with the second horn of the dilemma, aiming to find an appropriate theory that could possibly save mental causation. The questions that I would like address in this chapter is the following: what is it about epiphenomenalism that is so "repulsive"? Can we just accept the conclusion that our mental properties do not have causal powers? They exist, sure, but they are causally inefficacious.

The problem with epiphenomenalism is that such a view is highly implausible. Kim states immediately that epiphenomenalism is "obviously wrong, if not incoherent." (2005: p. 70) Interestingly enough, out of the vast literature on mental causation, there is no serious debate on validity of epiphenomenalism[9]. It is considered an eccentric position, usually outright rejected as an option, with authors directly trying to find a solution to the causal exclusion problem (or some other problem of mental causation). There are some (e.g. Heil 2013) who mention the unavoidable battery of experiments conducted by Benjamin Libet (especially his 1985 paper), trying to hint at a possibility of epiphenomenalism of being a *real* option. Nonetheless, it would still be philosophically questionable to claim that such experiments consequently corroborate epiphenomenalism.[10] As instead, epiphenomenalism tends to be viewed in the similar light as radical scepticism. On such a view, the purpose of epiphenomenalism would be to question common sense, a version of a quite intriguing "what-if…" I concur with Kim when he claims that:

[9] One of rare exceptions would be Frank Jackson.

[10] Libet's experiment shows that 550 ms before a voluntary act, a cerebral activity occurs, implying that "voluntary acts can be initiated by unconscious cerebral processes before conscious intention appears". However, Libet also resists absolute impossibility for volition to have no effect whatsoever, claiming that "conscious control over the actual motor performance of the acts remains possible", e.g. vetoing the actual motor activation. (1985: p. 529).

> Just as reflections on sceptical arguments have deepened our understanding of the nature and limits of human knowledge, "worries" about epiphenomenalism may well lead to a deeper understanding of just what our mentality consists in and how it is related to our physical nature. (1998: p. 62)

This seems to be true, as the epiphenomenal consequence of the causal exclusion argument has created a plethora of papers and books trying to unveil this mystery that is mind-body interaction; this one included.

One final remark before venturing onto different resolutions to Kim's exclusion problem: epiphenomenalism claims that the mental is causally inert, but this is very implausible. Kim says, "it is the problem of showing *how* mental causation is possible, not *whether* it is possible." (1998: p. 61) Bearing that in mind, let us see *how* reductive physicalists try to resolve the issue of causal exclusion. I will briefly analyse two options: *identity theory* (or *type physicalism*) and *eliminative materialism*.

2. REDUCTIVE PHYSICALISM

There are two radically opposing options that this chapter will show, and their common denominator will be *reductionism*, the idea that mental states can be reduced to physical states. One option is that we might be able to save mental causation and claim that mental properties can have causal powers, but only if they are identified with physical states. The other option is slightly more radical, as it implies that not only are mental states reducible, they can also be *eliminated.*

2.1. Type Identity Physicalism

During the 1950s, a series of papers had been published which would radically change the direction of philosophy. Namely, papers published by Herbert Feigl, U. T. Place, and J. J. C. Smart all shared one common thread: the idea that "minds are material entities – brains – and mental properties are, as a matter of discoverable empirical fact, material properties of brains and nervous systems." (Heil 2013: p. 74; also Berčić 2012: p. 163). In the first part of this paragraph I would try to unbox this idea and see what it really implies.

First and foremost, the motivation for such an undertaking comes from scientific development of that time. It was first believed that *correlations* had been located between certain mental and physical states. The most famous of such correlations that inspired this entire debate has been pairing up "pain" with the "activation of C-fibres". Now, let us suppose that for each occurrence of a certain mental property we have successfully paired it up with its physical *correlate.* The question is, what do we make of this correlation? Surely, it cannot be just an accidental co-occurrence. There must be something to it. What Smart and other mentioned philosophers argued was it is something more than correlation; the mental states *are identified with* the physical states. In the same way as we have identified lightning with electromagnetic discharge, or temperature with mean molecular kinetic energy, or water with molecular structure of H_20, we have to identify pain with the activation of C-fibres.[11] Smart concludes that:

> should [visual, auditory, and tactual sensations, aches and pains] be *correlated* with brain processes does not help, for to say that they are *correlated* is to say

[11] It does have to be noted that in the following years, the exact neurosynaptic basis for pain was shown to be a bit more complex than just the activation of C-fibres, but I do not think that this completely changes the discussion.

> that they are something "over and above". You cannot correlate something with itself." (1959; p. 142 emphasis unchanged)[12]

A more general principle of the identity theory is provided by Heil:

(I) a state, event, or process, α, is identical with state, event, or process β, only if the properties involved in α and β are identical. (2013: p. 75)

To gain a better understanding of this principles, let us use these two propositions:

a) Superman is a superhero.
b) Superman is Clark Kent.

It seems that in both these cases we have used some kind of identification. In the first case, we have identified Superman as a superhero, and in the other we have identified Superman as Clark Kent (and both these propositions are true). However, only the second proposition really makes a statement of α is β. The first proposition is one of *predication;* we have provided an attribute to Superman, it says something *about* him, but it is not the statement of identification (as there are countless superheroes nowadays)[13]. So, analogous to identifying pain with C-fibres; we haven't simply provided an attribute, or a characteristic of pain. No, we have *identified* it with some brain process, we have shown what it actually *is*, same as we have shown what lightning or temperature actually *is.*

The question that we pose ourselves now is: how does this help us with the problem of mental causation? The solution is simple. Mental properties indeed *do* have causal powers and they enter causal relations *without* breaking the principle of physical closure because mental properties just *are* physical properties (they are nothing *over-and-above* physical properties). For example, if we claimed that pain caused us to wince and groan, and pain is nothing more than the activation of C-fibres, then we (i) don't encounter the problem of overdetermination since there are not two competing causes and (ii) since C-fibres caused us groan and wince, we thus also avoiding the problem of physical closure. Clearly, this isn't something that many philosophers would like to have, as it immediately implies at least *some sort of* epiphenomenalism; it is not pain *in and of itself* that caused me to groan and wince, and I believe that it really was the pain, not some C-fibres. Pain once again seems to be irrelevant.

[12] This is also why this version of identity theory is also called *type identity*, as one type of thing is reduced to another; pain as a type is reduced to and identified with C-fibres, a physical type.
[13]This distinction is important when it comes to notions of *numerical* and *qualitative* identity. "It's the same hat" can mean that it is numerically one and the same hat, or it can mean that it completely resembles another hat.

However, I do believe that such a view simply does not fully grasp the idea behind the identity theory. The identity theory does not claim that pain has no causal powers, quite contrary, it does. Pain causes us to groan and wince insofar as *it just is* the activation of C-fibres.

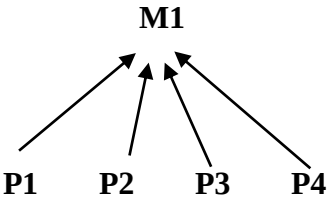
Such a view certainly has a lot of merits. First, scientific advancements go hand in hand with philosophical theory. Naturally, the sciences and especially neurosciences, still have a lot of unpacking and understanding to do. Although we are now closer than ever to understanding, deciphering, or mapping out our brain, there are still countless phenomena that we have yet to explain. It is just a matter of time when we succeed in doing that. One can often encounter arguments (albeit amongst scientific and philosophical laymen) that medicine or sciences in general have still not found answers to some questions. Therefore, we will never be able to completely reduce all of our mental states to corresponding physical ones. I believe such an inference is not valid. Nevertheless, it gives motivation to scientists to persevere in finding answers. Second merit that I would like to point out is that philosophically, and argumentatively, identity theory makes a strong case when it comes to saving mental causation without breaking principles of overdetermination and physical closure. However, as I am going to show now, physicalism may have a downside, which would lead us into rejecting it, at least in the form of identity theory, and pursue us into finding another option.

2.2. Functionalism and Multiple Realizability

Despite gaining momentum in the 1950s, identity theory in this form soon came to its downfall. The counterargument directed towards the physicalism was rather simple, but to a large extent put an end to identity theory, at least in this form. The argument is the so-called *argument of multiple realizability*, as famously suggested by Hilary Putnam. What Smart and other physicalists wanted was one-to-one identification of mental states with physical states. However, according to Putnam we can easily imagine some non-human beings that would arguably share the same mental states, but instead have something else activated. For example, we can imagine a Martian who is feeling pain and instead of C-fibres, in his case M-fibres are activated. The question is, how is it possible that they share completely same mental state, but they have different physical ones? According to Putnam, the main problem is that we are going about mental states in a completely wrong manner. This serves as the basis for *functionalism*.[14]

[14] The most notable figures when it comes to functionalism are Hilary Putnam and Jerry Fodor.

Although it is difficult to find a unitary definition of functionalism, the key notion essential to functionalism are *functional roles* that our mental states have, as opposed to trying to reduce them to their neurophysiological basis.[15] So, according to this definition, we should observe pain, for example, as a mental state with specific functional roles (usually its role is to make us groan, wince, or something else). (Heil 2013) The problem is not that we *cannot* find the physical base, the problem is that different *physical realizers* can realize one mental state (best depicted in Figure 1), consequently disabling us from positing type identity.



(the mental state, e.g. pain)

(physical realizers, e.g. C-fibres, M-fibres, etc.) (Figure 1)[16]

Furthermore, it is useful to think of these relations in terms of *levels*; pain being a *higher-level* property, and C-fibres, M-fibres, or all other possible pain's *realizers*, being *lower-level* properties.

Based on these arguments, there are two conclusions that can be inferred from Putnam's argument: (i) type identity as required by identity theorists is not sustainable and (ii) mental states might not even be reducible to physical states. Whether (ii) is true or not still remains to be seen, but according to the functionalism that Putnam suggest, this seems to be the case. Putnam's legacy is still present in some *non-reductive* approaches to mental causation, as we will see in the chapter 3. We have seen how one reductive option fails to save mental causation. Can the second option, one of eliminative materialism save it?

2.3. Eliminative Materialism

If identity theory, or type physicalism, was at the one end of the reductive spectrum, this theory is on the other; and that is *eliminative materialism*. One of the most famous

[15] For these arguments I have used two sources; Berčić (2012) and Heil (2013). Putnam's paper that I have used was *The Nature of Mental States* from his *Mind, Language, and Reality*: *Philosophical Papers*, vol. 2 (1975). That paper is a reprint of his 1967 paper titled *Psychological Predicates.*

[16] This figure also shows the relation of levels from the chapter 3.1., with the mental states being higher-level properties, and their realizers lower-level properties.

proponents of this position, Paul Churchland[17], opens his 1981 paper on eliminative materialism with the following strong claim:

> Eliminative materialism is the thesis that our commonsense conception of psychological phenomena constitutes a radically false theory, a theory so fundamentally defective that both the principles and the ontology of that theory will eventually be displaced, rather than smoothly reduced, by completed neuroscience. (1981: p. 67)

The idea is quite radical, and it requires of us to completely abandon any ontology that contains our mental states. The theory is originally a critique of the *folk psychology*, considering it to be completely *false* and *wrong*.[18] This critique consequently shows that we should reduce any talk of mental states to talk of neurophysiological processes. This reduction, in opposition to physicalism presented in the last chapter, *eliminates* any and all mental states. To get a clearer picture of how eliminativism of this kind should work, we just need to observe the history of science; before we believed in phlogiston, but then it was replaced by a better theory of oxidation, and thus phlogiston was eliminated. Similarly, before scientists believed in *vital spirit*, but then it was proved to be incorrect and false, and thus completely eliminated. Now, argues Churchland, folk-psychology is incorrect and false, and thus we should completely reject the ontology of mental states, and replace it with neurosciences.

What makes this position interesting in the context of this paper is how it actually deals with the problem of mental causation. Because once you eliminate mental states, you consequently eliminate all issues relating to mind-body interaction. For that reason, there is no problem of mental causation anymore, because there are no mental states. When we say that I ordered a salad instead of a chocolate cake because I *thought* it would be the best option for me and my health, such an interpretation is utterly wrong because there are no such things as my intentional mental states. Instead, such talk of what caused me to choose a salad over a chocolate cake should be reduced to neurosynaptic explanations; probably somewhere in my prefrontal cortex a network of neurological pathways activated which led me to taking a salad.

This is certainly something appealing in this theory. However, it is far from tenable. There are different ways to refute eliminative materialism (Lycan 2005), from it being *self-refuting*, i.e. an eliminative materialist has to claim that they *believe* that there are no beliefs,

[17] Together with his wife Patricia Churchland, they have written numerous papers and books on this topic. Another proponent of this theory is Stephen Stitch, but in this work, I will only use the paper by Paul Churchland.
[18] A great outline of the arguments proposed by eliminative mater is set by Lycan (2005).

to simply rejecting it on the common-sense basis. In my opinion, a much stronger argument against eliminativism is that although folk psychology may be imprecise, it is far from being completely wrong, as we tend to have successful inferences about our behaviour based on our mental states. Not only that, but if there are no mental states whatsoever, how is it possible that we were at least somewhat successful in predicting and explaining certain behaviours?[19] I would not like to get far into details here. As instead, I would like to draw the same conclusion as I did with epiphenomenalism. I believe that eliminativism has interesting and thought-provoking upshots that motivate further inquiry in philosophy of mind. When it comes to any discussion on mental causation and its relation to folk-psychology or neurosciences, there are still a lot of unclarities that are left to uncover. But if we simply *eliminate* such possibility to move forward, we will never make any progress. As instead, eliminativism should serve more as a motivation, rather than a solution to our problems.

In the next chapter I turn to non-reductive approach to saving mental causation. Though, I have to mention one thing: identity theory and eliminativism are not the only reductive models trying to save mental causation. In chapter 6, I will analyse Kim's solution to the problem of causal exclusion, which is going to functional reductionism.

[19] See Biondić (2017).

3. NON-REDUCTIVE PHYSICALISM

Going to back to the horns of our initial dilemma, in this chapter I would like to answer the following question: how do we save causal powers of mental states in a physical world, *without* having to reduce them neurological ones? Surely, there must be something more to our mental states other than neurosynaptic and physiological processes? But how do we account for that distinctness? I will analyse two ways that account for this distinctness: non-reductive physicalism defined in more general terms and Davidson's anomalous monism. Although there is no unified or a single theory of non-reductive physicalism, there are perhaps three doctrines that all variances of this theory share, at least according to Kim (2005), and they are: (i) *mind-body supervenience*, (ii) *the physical irreducibility of the mental*, and (iii) *the causal efficaciousness of the mental*[20]. What these principles aim to show is that mental properties have causal powers by being supervenient on physical properties. According to this argument, any time a physical property upon which a mental property supervenes causes another property, that mental property also causes it. I believe that this approach to mental causation is false and in the next chapter I will argue that it is false on the basis of the causal exclusion argument.

3.1. Three Principles of Non-reductive Physicalism

The concept of supervenience implies a hierarchy of properties, where higher-level properties are determined by, or depend on lower-level properties. Kim defines it as:

> *Supervenience*. Mental properties strongly supervene on physical/ biological properties. That is, if any system *s* instantiates a mental property M at *t*, there necessarily exists a physical property P such that *s* instantiates P at *t*, and necessarily anything instantiating P at any time instantiates M at that time. (Kim 2005: p. 33)

Kim has probably been the most influential philosopher in developing the concept of supervenience (Kim 1993), which one might find ironic, as Kim is going to criticise the same concept that he refined. What does supervenience entail?[21] Let us take the already used example of pain and C-fibres. Supervenience suggests that when a person feels pain (*s* instantiating a

[20] The analysis that follows is based solely on Kim's account of non-reductive physicalism (2005: p:33-5). Although it is possible to find nuances in definitions, I believe that every non-reductive physicalist would be able to accepts these doctrines.

[21] The concept of supervenience occurs in different contexts and its origin is still up for debate. It is used in different areas of philosophy, for example aesthetics, ethics, and other fields. For the current purposes, I am using supervenience to depict relation between mental and physical states.

mental property M), that person is also instantiating a physical property at the same time, which is in this case the activation of C-fibres. And, further on, every time a person's C fibres are activated (instantiating P), that person also feels pain (instantiating M). The main question is: how might supervenience help non-reductive physicalists in saving mental causation? As Kim argued in 1993: "when a mental event M causes a physical event P, this is so because M is supervenient upon a physical event, P*, and P* causes P." (Kim 1993: 106) In other words, when my pain is causing me to groan or wince, it is doing so by being supervenient upon C-fibres, and C-fibres are causing me to behave in that manner.

One thing is worth mentioning: supervenience does not entail mere correlation; that every time C-fibres activate, pain follows. It is a stronger, metaphysical claim of dependence, the instantiation of pain occurs *because* C-fibres get fired up; supervenience "includes a claim of existential dependence of the mental on the physical." (Kim 2005: p. 34). Just to avoid any further confusion, one note before moving on to the next doctrine. Supervenience is an asymmetrical relation, meaning that if the mental supervenes on the physical, it cannot be the other way around. (Kim 1984: p. 67) Another useful interpretation of this relation would be to say that every change in the supervenient property (in this case that would be mental property, or pain), implies changes in the *base* property (physical property, or C-fibres).[22]

Crucial to the non-reductive physicalism is the second doctrine and it states the following:

> *Irreducibility*. Mental properties are not reducible to, and are not identical with, physical properties. (Kim 2005: p. 34)

This doctrine is first and foremost a reaction to Smart and other identity theorists that were analysed in the previous chapter, and a continuation on functionalism. What non-reductive physicalists claim is that there are certain properties outside the domain of physical. For example, pain may be supervenient on some physical properties (activation of C-fibres), but it is certainly not reduced, or identical to them. This move is motivated by the argument of multiple realizability, according to which mental properties (or types, to be more precise) are not reducible to physical, as there can be different physical properties that realize the same mental property. Another notion of irreducibility will be encountered soon in Davidson's

[22] For this interpretation of supervenience, I am using Kim's *Concepts of Supervenience* (1984).

anomalous monism, which together with functionalism are "the best-known examples of non-reductive physicalism." (Kim 2005: 158)

Finally, such mental properties which are irreducible and supervenient on physical properties are, according to these theorists, *causally efficacious.*

> *Causal efficacy*. Mental properties have causal efficacy—that is, their instantiations can, and do, cause other properties, both mental and physical, to be instantiated. (Kim 2005: p. 35)

I believe there is no need to once again elaborate on this doctrine, as it is the first premise from which our arguments begin. For example, pain that we experience causes us to groan and wince (physical effects), and it can also cause some emotional distress (mental effects).

3.2. Anomalous Monism

One of the pioneers in the non-reductive approach to physicalist interpretation of mind-body relation was Donald Davidson, with his influential paper *Mental Events* (1970), where he proposed a new theory of the relation between the mental and the physical; *anomalous monism.* Now, although it is possible to write a whole dissertation on just one of the tenets and principles of his position, I will provide just a brief summary of his position. This elaboration on Davidson's paper will show how his theory relates to a more general approach to non-reductive physicalism depicted above, and how both of these positions generate what Kim calls the causal exclusion problem.

First, what is Davidson's aim? Davidson starts off his essay with a quote on Kant who tries to reconcile "freedom and natural necessity in the same human action" (cf. Davidson 1970: 207). Similar motivation strikes Davidson, as he is finding a way to reconcile causal role of mental events without having to completely reduce them to physical events, and he tries to do that by supposing *anomalous monism.* So, the best way to understand Davidson, in my opinion, would be to simply dissect what *anomalism* and *monism* stand for in his theory. When it comes to anomalism, Davidson starts off with three principles which seem to stand in an 'apparent contradiction' (1970: p. 208).[23] They are the following[24]:

[23] This 'apparent contradiction' is a reference to Immanuel Kant

[24] Together with Davidson's paper (1970), two other sources will be used; Robb and Heil (2019) and Kim (1998). There are numerous other secondary sources that I have found useful in interpreting Davidson, for example Heil and Mele (1993), although here they refer mostly to his later work.

Principle of Causal Interaction: Some mental events interact causally with physical events. (Davidson 1970: 208)

I have already put forward different examples and cases of this. Davidson uses an example of sinking the Bismarck and events that played a causal role in bringing about sinking the Bismarck, "such as perceivings, notings, calculations, judgements, decisions, intentional actions, and changes of belief" (p. 208) Such cases don't seem problematic, at least not *prima facie.* The problem occurs when the next two principles are introduced.

Principle of the Nomological Character of Causality: Events related as cause and effect fall under strict laws.[25] (Davidson 1970: 208)

Anomalism of the Mental: There are no strict laws on the basis of which mental events can be predicted and explained. (Davidson 1970: 208)

Although Davidson initially doesn't offer much of elaboration on what is meant by the last principle, the gist is that unlike laws of physics which are deterministic (or exceptionless, as Robb and Heil (2019) interpret it), laws of psychology are not! This constitutes *anomalism*, and this is also the *non-reductive* element of Davidson's theory. According to Davidson, as there are no strict psychological laws like there are physical laws, mental events *are not reduced to* physical; pain is not reducible to some physical realizer (for example C-fibres) on the basis of absence of strict psychological laws. Davidson does claim that "mental characteristics are in some sense dependent, or *supervenient*, on physical characteristics." (1970: p. 214), something that we have seen with non-reductive physicalism in general, making it clearer in what sense Davidson is a non-reductive physicalist. However, if we accept this, we still need to find a way for mental events to enter causal relations. Since *the principle of causal interaction* requires strict laws, *the nomological character of causality*, and the mental lacks it. This is the apparent contradiction that Davidson anticipated. However, Davidson has a way out.

Davidson's ingenious solution is simply to say that every mental event is in fact also a physical event – hence the *monism*! (1970: ch. 2) Kim perfectly emphasises this monism in the following paragraph:

> (…) mental events in causal relations must instantiate laws but since there aren't any psychological laws, that can only mean that they instantiate physical laws. This shows that mental events fall under physical kinds (or have true physical

[25] In other words, where there is causality, there must be a law. (Davidson 1970: p. 208).

> descriptions), from which it further follows, argues Davidson, that they are physical events. (1998: p. 33)

Therefore, taking all Davidson's principles into consideration, I believe that Davidson tries to save mental causation insofar as mental events indeed enter causal relations, but they instantiate physical laws since there are no psychological strict laws, thus claiming that those mental events are also physical events, although certainly not reducible to (but perhaps supervenient on) them.[26] In Davidson's own words:

> Anomalous monism resembles materialism in its claim that all events are physical, but rejects the thesis, usually considered essential to materialism, that mental phenomena can be given purely physical explanations. Anomalous monism shows an ontological bias only in that it allows the possibility that not all events are mental, while insisting that all events are physical. (Davidson 1970: 214)

Although Davidson differs from functionalists in some aspect, the key notions of non-reductive physicalism are still present, especially mental efficacy and irreducibility. Supervenience was developed later on, but Davidson's argument saves this notion of having both mental and physical, without reducing one to another *and* saves the notion of mental causation.

However, is this enough? I do not think that these conditions will suffice for preserving mental causation and in the following chapter I will show why. For now, let us see a parallel that Kim draws between Descartes and Davidson:

> there is an instructive parallel between Descartes's mind-body problem and the way the current debate on mind-body causation arose. For it was Davidson's anomalous monism that first touched off the current worries about mental causation. ... In a way that is reminiscent of Descartes's sharp separation of mind from matter, Davidson conceived of mental phenomena as constitutively distinct from physical phenomena… And yet, like Descartes, Davidson wanted causal interaction between mental and physical events… The only difference between Descartes and Davidson is that for the former it was the dualism of substances that caused the trouble, while for the latter it was his dualism of properties. (Kim 1998: p. 58)

[26] A brief note; Davidson's interpretation here is one of *token* identity (in contrast to *type* identity we have seen before). On the example of pain, this only means that individual cases of pain are also individual cases of *some* physical kind.

4. CAUSAL EXCLUSION AND NON-REDUCTIVISM

Although non-reductive physicalism strived towards saving the intuition of both having mental and physical *without* reducing one to another and without breaking the principle of physical closure, there was still one major problem that followed – *the problem of causal exclusion*. The problem is a result of inconsistency between the principles of non-reductive physicalism and the principles of physical closure and overdetermination. I have already anticipated the argument in a simpler form and now I will present it step by step, as according to Kim (2005). To list them once more for clarity, these are the principles that will render inconsistency and show that non-reductive physicalism cannot defend mental causation: *Mental Efficacy, Supervenience, Irreducibility* – these three being principles of non-reductive physicalism, and *Physical Closure* and *Overdetermination.*

First, I would like to present one of the simpler versions of the argument. For each line of the argument I have added which principle it stands for, although the exact and complete form of the argument is by Yablo (1992):

> If a property F is causally sufficient for a property G, then no property distinct from F is causally relevant to G, barring overdetermination." (*exclusion – overdetermination)*
> For every physical property P*, there is a physical property P that is causally sufficient for P. (*physical closure)*
> For every mental property M, M is distinct from P. (*irreducibility)*
>
> ---
>
> For every physical property P*, there is no mental property M that is causally sufficient for P*. (*exclusion, i.e. epiphenomenalism)* (Yablo 1992, 247-248)

What follows is the more detailed version of the argument, given by Kim. To completely understand the argument, I shall now provide a detailed step-by-step analysis. Each line of the argument is taken from Kim, accompanied by explanation of what they present and how one line follows from another. To say once more, in his argument Kim assumes aforementioned five principles and wants to show that they yield inconsistency!

First, let us assume a case of mental-to-mental causation. The argument will work for both mental-to-mental and mental-to-physical. For example, imagine a case where pain causes us to groan and wince (mental-to-physical), but it also causes us some distress (mental-to-mental). This is going to be the doctrine of *Mental Efficacy*:

(1) M causes M*. (Kim 2005: 39)

From *Supervenience* it follows that:

(2) For some physical property P*; M* has P* as its supervenience base. (Kim 2005: 39)

Now, P* necessitates the occurrence of M*, which begs two options; one is that P* was already instantiated, without M, or M had something to do with the occurrence of P*. In other words, in order for M* to be instantiated, P* has to be instantiated. Therefore, M* first has to cause P* in order to bring about M*. Therefore:

(3) M caused M* by causing its supervenience base P*. (Kim 2005: 40)

These three steps so far show that if we want supervenience to work, in order to have mental-to-mental causation, we first need to have mental-to-physical. Furthermore, from supervenience it also follows that:

(4) M has a physical supervenience base, P. (Kim 2005: 41)

Here's where it gets tricky. From *Closure* and from (3) it follows:

(5) M causes P*, and P causes P*.[27] (Kim 2005: 41)

Further on, from *Irreducibility* it follows:

(6) M ≠ P (Kim 2005: 42)

From (5) and (6) it follows that:

(7) P* has two distinct causes, M and P, and this is not a case of causal overdetermination. (Kim 2005: 44)

This step leads us to the causal exclusion principle, which states that: "No single event can have more than one sufficient cause occurring at any given time—unless it is a genuine case of causal overdetermination." (Kim 2005: p. 42) This leaves us with two options, either we have to let go of M as a cause or P. For Kim, the solution is simple:

(8) By *Closure* and *Exclusion*, M must go; P stays. The putative mental cause, M, is excluded by the physical cause, P. That is, P, not M, is a cause of P*. (Kim 2005: 44)

[27] Another version of this step would be "P* has a physical cause—call it P—occurring at the time M occurs." (Kim 2005: p. 44)

Just to elaborate on this a bit further. Why not choose M as the cause of P* instead of P? Well, if we wanted that, we would break the principle of *Closure*, as there must be a physical cause of P*. Second, why can't we say that *both* M and P caused P*? The problem lies in *Overdetermination.* Overdetermination works only if there are two separate, independent causal chains that lead to the same effect. However, this is not the case because of *Supervenience.* Overdetermination claims that we can imagine a possible world in which one causal chain hadn't occurred, but the effect is still the same because of the other causal chain. With mental causation that is not the case, as we cannot imagine a possible world in which mental chain does not occur without its *subvenient* physical base. Thus, those two chains are not sufficient, as one depends on the other.[28] As I final touch to this argument, I shall provide a figure (Figure 2) that perfectly encapsulates the argument.[29]

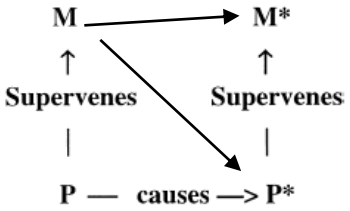


Figure 2.

This brings us back to the same upshot from the beginning of the paper: *epiphenomenalism.* Kim clearly finds this conclusion hard to accept, stating that "the idea that our thoughts, wants, and intentions might lack causal efficacy of any kind is deeply troubling, going as it does against everything we believe about ourselves as agents and cognizers." (2005: p. 70) We can reject epiphenomenalism, but that still does not resolve the question of how exactly mental properties have causal powers.

One of the ways in which non-reductive physicalists tried to battle off Kim's argument of causal exclusion was by appealing to counterfactual analysis of causation, most famously developed by David Lewis. In the next chapter I will provide an analysis of that approach, as well as a brief introduction to Lewisian counterfactuals. Finally, I will show why such an approach is not sufficient for preservation of mental causation.

[28] These arguments are presented in Kim's chapter *Is Overdetermination and option?* (2005: p. 46-52)
[29] Barry Loewer calls this "Kim's favorite diagram" (2007: p. 257). Similar diagram is also showcased by Sesardić (1984)

5. A COUNTERFACTUAL ACCOUNT

The central idea of mental causation is that physical effects *depend on* mental states. This notion of dependency, argue some philosophers, is enough to claim that mental properties can indeed be *difference makers.* Some authors believe that this dependence is *counterfactual* dependence and that *counterfactual conditionals* are the best way to grasp it. Tyler Burge (1993), for example, argues:

> one can specify various ways in which mental causes 'make a difference' which do not conflict with physical explanations. The differences they make are specified by psychological causal explanations, and by *counterfactuals* associated with these explanations. (p. 115; emphasis added)

The notion of counterfactuals is important to non-reductive physicalism and it aims to argue the following: had the mental property not occurred, the physical effect would not have occurred. The idea of counterfactuals is rather complex, but I will try to briefly explain it before getting to the core of counterfactual mental causation.

5.1. Causation and Lewisian counterfactuals

Lewis' counterfactual analysis of causation is a multi-layered, complex theory that's been in development for decades now and it would be ungrateful towards Lewis to sum it up in less than a page. However, we have to briefly analyse Lewisian idea of counterfactuals, since they play a huge role in contemporary philosophy, especially metaphysics. Despite its complexity, in the crux of this theory, there is a seemingly simple proposition. When we claim that an event *c* caused event *e*, by that we mean that if *c* had not occurred, *e* would not have occurred. (Lewis, 1973) Let us take an example, and see what this actually implies. Suppose that a striking of a match caused it to light. Counterfactual analysis of this claim would result in the following proposition:

C – If the match had not been struck, it would not have lighted. (Kim 2007: 233)

So far, this seems to be unproblematic. However, what is it that makes this proposition to be true? Here is where it gets slightly complicated. In order to see what makes this proposition true, we need what is called *comparative over-all similarity* among possible worlds. (Lewis 1973: p. 8-13) The main line of argumentation is the following: out of two possible worlds, the

closest one is that which bears the most similarity to our actual world, except for one thing: in that world the match had not been struck. So, we need to observe what happens in that world in which the match had not been struck. If in that world the match does not light, then C is true. The problem is, how do we know that this world is the closest, and not the world in which the match had not been struck, but the match nonetheless lighted? Kim explains this in the following paragraph:

> The obvious, and the only possible, answer seems to be that, in the actual world, dry matches struck in the presence of oxygen usually and reliably ignite, and that it is our knowledge of this regularity, or law, combined with knowledge of the actual circumstances in which the match was struck (e.g., it was dry, oxygen was present, etc.), that accounts for the judgment that the world in which the match that was not struck does not light is closer to the actual world than is the world in which the unstruck match lights. (2007: p. 234)

5.2. Counterfactual account of mental causation

Now, let us see how this analysis applies to mental causation. Kim considers a causal claim that "a sudden attack of migraine headache caused Susan a frightful sense of anxiety" (2007: p. 234). Following the counterfactual analysis similar to last paragraph, the proposition we are looking for would amount to:

D – If Susan had not had the sudden migraine headache, she would not have experienced frightful anxiety. (Kim 2007: p. 234)

This seems reasonably acceptable, as we tend to claim similar propositions on almost daily basis. Nonetheless, we need to observe the closest possible world to ours where everything remains the same but the fact that Susan had not had the sudden migraine headache. If in that world Susan had not had a headache, and *did not* experience frightful anxiety, that would make *D* true. However, similarly to the example of matches in the paragraph, we need to ask ourselves the same question; what is it in this occurrence that makes it in fact true? C was true on the basis of regularity, or law, in combination with actual circumstances. So, can the same be applied to *D*? LePore and Loewer (1987) argue that it can. They believe that one of the faults of Davidson's anomalous monism (and mental causation in general) is that it required *strict laws*. Instead as LePore and Loewer claim, "a non-strict law may be improved upon by

explicitly including some of its *ceteris paribus*[30] conditions in its antecedents." (LePore and Loewer 1987: 632) Or, another option would be to opt for some kind of regularity. Kim elaborates this idea by claiming that the truth of *D* in that case must depend:

> on the regularity connecting sudden attacks of migraine headaches and feelings of anxiety. This regularity could be limited to Susan and others like her in relevant (presumably neurophysiological) respects, or it could be a (*ceteris paribus*) law for all people with migraine headaches. (2007: p. 234)

Is this enough to save mental causation without having to reduce mental to physical? I believe that LePore and Loewer, together with Loewer's later work (2002; 2007) made a really strong case for saving mental causation in non-reductionist spirit. The notion of causal relevance, or dependency surely keeps the intuitive sense of importance of our mental properties in the causal chain leading to desirable effect. However, is such an approach sufficient, or we still need something more?

5.3. Two problems of counterfactual (mental) causation

I believe that the approach depicted above has a lot of merits. However, I do not think that counterfactual approach to mental causation can save it. There are two issues I find with this approach that I would like to elaborate on. The first one can be traced back to epiphenomenalism. Even a supporter of epiphenomenalism would agree with the statements "had not M occurred, P would not have occurred", but this does not warrant a causal relation between M and P. The best way to describe this would be to use the example of matches. We have seen that the counterfactual claim "if the match hadn't been struck, it would not have lighted" is true, but let us also add another effect from striking match. For example, if the match hadn't been struck, smoke would not have appeared. So, an event *c* (striking a match) has two effects; e_1 (smoke – an effect which can be epiphenomenal) and e_2 (lighting of the match). According to counterfactual analysis, a proposition "if e_1 had not occurred, e_2 would not have occurred" is in fact true, but that does not mean that e_1 caused e_2. In this case, the counterfactual would be: if the smoke hadn't appeared, the match would not have lighted.

This is only one of many different objections and obstacles that counterfactual analysis of causation has come across in the past couple of decades. Naturally, counterfactualists have

[30] *Ceteris paribus* means all things remaining unchanged.

been battling off these objections (together with the objections of *preemption*, *overdetermination*, *context dependence*), and to some extent quite successfully. But the question that I would like to pose still remains; should we base our idea of mental causation on a causal theory that is still very much under question?[31] I do believe that appealing to counterfactuals is quite useful, especially when it comes to mental causation; it is a useful tool in understanding regularities, laws, and perhaps even in providing some sort of explanation. But when it comes to tracking causal processes, we require something more, something that *generates* or *produces* physical effects. This brings us to the second objection.

Kim quotes Ned Hall (2004) in order to explain distinction between generative causation and causation as dependence (or relevance):

> Causation, understood as a relation between events, comes in at least two basic and fundamentally different varieties. One of these, which I call "dependence," is simply that: counterfactual dependence between wholly distinct events. In this sense, event *c* is a cause of (distinct) event *e* just in case *e* depends on *c*. That is, just in case, had *c* not occurred, *e* would not have occurred. The second variety is rather more difficult to characterize, but we evoke it when we say of event *c* that it helps to *generate* or *bring about* or *produce* another event *e*, and for that reason I call it "production." (Hall, 2004, p. 225)

Based on this distinction, the counterargument to counterfactual analysis of (mental) causation would be the following: counterfactual approach to mental causation is not strong enough because it does not grasp generative or productive element of causation. In my opinion, the generative or productive element of causation is more important than the counterfactual one. The main reason why we would like to have generative notion of causation can be traced back to the second chapter, where I have argued that the main motivation for this whole discussion lies in saving the idea of human agency. Now, although Loewer and similar philosophers might argue that productive notion of causation is too strict[32], it nonetheless seems to be a condition for human agency. The mere idea that my desires *caused* me to behave should perhaps be interpreted in the most straightforward sense; not only did my action *counterfactually depend* on my desires, but my desires *generated*, or *brought about* desired actions. I concur with Kim when he claims that "mere counterfactual dependence is not enough to sustain the causal

[31] My bachelor's thesis revolved around counterfactuals, where I have presented and put forward the arguments of preemption, overdetermination and context dependence, so for that reason I shall not be getting too much into details, as instead I will keep my focus on mental causation specifically. (Kobašlić 2016)
[32] Loewer might even argue that it does not deserve a place in contemporary physics (2002).

relation involved in our idea of acting upon the natural course of events and bringing about changes so as to actualize what we desire and intend." (2007: p. 236)

One objection that someone might point towards this argument would be to say that generative causation completely negates the possibility of mental causation in this straightforward sense. If it were truly possible, we would break the principle of physical closure. Have we once again reached the epiphenomenal conundrum? I do not think we have. In the final chapter of this thesis, I would like to present an option according to which we can have generative mental causation, but non-reductive physicalists are probably not going to like it. In brief, this is the argument: the only way we can save mental causation in terms of generating desired effects is to (i) find physical realizers of mental properties that enter causal relations, and (ii) reductively identify mental properties with their physical realizers.

6. FUNCTIONAL REDUCTIONISM

6.1. Can reductionism save mental causation?

At the beginning of the chapter 4 of his *Mind in a Physical World* (1998), Kim depicts a current attitude towards reductionism in general. Reductionism is often frowned upon, "it goes beyond mere criticism or expression of doctrinal disagreement; it is to put a person down, to heap scorn on him and his work." (Kim 1998: p. 89) I don't mean to provide an apology for reductionism in this chapter, but I do want to paint a picture according to which reductionism should not be viewed as a downfall of philosophical belief. As instead, some of the notions of reductionism should be clarified and we should cast some light upon certain principles of reductionism, as it could truly help us in gaining better understanding of the phenomenon that is in question – mental causation and mind-body interaction. The task is not easy. I would like to point out one argument from Kim (2005), in which he claims the following:

> (…) even if we are entitled to the conditional "Mental causation is possible only if some form of reductionism holds," we cannot infer that reductionism is true. For we should not a priori rule out epiphenomenalism and other noncausal views of the mind. Reductionism must be earned—and mental causation, too, must be earned—by showing that mental properties are indeed reducible, in a relevant sense, to physical/biological properties. (p. 148)

This attitude towards reductionism was already briefly depicted when we discussed the identity theory and eliminative materialism, as both of these options are reductionist. So, if reductionism is not an option, what do we have left? Our original dilemma posed the following options: either mental properties are epiphenomenal, or they have causal powers. If they have causal powers, they are either reductive or non-reductive. I do not think that they can have causal powers in non-reductive sense without facing the problem of causal exclusion. I believe that the only way to save mental causation would be to accept reductionism. Therefore, I would like to present one last model that just might save mental causation – Kim's *functional reductionism*.

6.2. Kim's model

This is going to be the last option that I am going to analyse, as it is also going to be one I personally endorse. Right from the start, *functional reductionism* might seem like an oxymoron; in the chapter 2.2. I presented functionalism as an opposing force to the identity

theory. Functionalists wanted to show that (i) mental states should be identified with functional roles that they perform and (ii) they can be multiply realizable. As we have seen, (ii) is a serious obstacle to the reduction of mental states to physical ones. In turn, functionalism, by being a position of non-reductive physicalism, had another problem and that is causal exclusion. So, the question is whether functional reductionism is a consistent option at all?

There are two things that I would like to point out before we proceed to Kim's model of reduction. The first one is that we have to distinguish two forms of functionalism; role functionalism and realizer functionalism. We have seen that Putnam belongs to the former group – the idea that mental states are to be identified with functional roles, whereas the latter form puts focus on typical physical realizers of mental states.[33] Kim provides his own model of reduction on the basis of realizer functionalism. The second important thing to observe is that Kim's model is a reaction to Nagel's model according to which it would not be possible to reduce mental to physical.[34] Kim argues against Nagel's model by claiming that "a Nagelian reduction of psychology to neurophysiology is simply irrelevant to the issue of reductively explaining psychological phenomena and laws on the basis of neurophysiological laws." (2005: p. 101) Therefore, Kim provides a three-step-plan on basis of which it would be possible to reduce mental to physical and thus save mental from being causally inefficacious. One of the main motivations as to why I am personally opting for Kim's view is because it stands out in the sense of keeping both mental causation and scientific implications of reduction, *without* having to sacrifice the mental. What remains to be seen is whether it falls into the trap of multiple realizability.

Kim's three-step model of functional reduction is the following. First, we take a property that we want to reduce, and we *functionalize* it; i.e. we define it in terms of causal tasks and role that they usually perform. Let us take into consideration two examples; pain as a mental property, and genes. Pain might be defined as a property caused by some injury, that usually performs a causal task of groaning or wincing. When it comes to genes, Kim provides the following definition: "a gene may be defined as being a mechanism that encodes and transmits genetic information. (2005: p. 101) After we have functionally defined the property,

[33] Proponents of this form are most notably Lewis and Armstrong. Heil (2013) provides definitions and distinctions of these forms, as well as Milojević (2017).

[34] Three criteria, or questions, are posited by Nagel's model of reduction, the so-called "bridge-laws", that pose problems to reducing mental to physical, and those are: the availability question, the explanatory question, and the ontological question, all three of which are analysed by Kim in (1998 and 2005). For the purposes of this thesis, I would not want to get into details of each of the criteria. Instead, I would like to focus on Kim's argument as it is more relevant to mental causation and causal exclusion as such.

the next step would be to find the realizers of such property. To be more precise, in terms of *higher- and lower-level* properties, we are looking for a lower-level property that completes the causal work of the higher-level property. So, when it comes to pain, that would probably the C-fibres (although, as I've mentioned before, the exact picture is a bit more complex for that, but for the purpose of this argument this will suffice). For genes: "it turns out that DNA molecules are the mechanisms that perform the task of coding and transmitting genetic information—at least, in terrestrial organisms." (p. 101) The third and final step would be to provide the exact explanation of the mechanism, i.e. how the realizers perform said causal tasks. For pain, we would require the explanation in terms or neurosynaptic activations, brain modules that interpret said neurons, and send back information to our hand which we move, or some other process that ends up in us groaning. "In the case of the gene and the DNA molecules, presumably molecular biology is in charge of providing the desired explanations", argues Kim. (p. 101). Once we have successfully performed all three steps, we have also functionally reduced (and identified) higher-level properties to their lower-level realizers. More formally, the model has the following form:

> Step 1 [Functionalization of the target property] - Property M to be reduced is given a *functional definition* of the following form:
> Having M = $_{\text{def.}}$ having some property or other P (in the reduction base domain) such that P performs causal task C.[35]
>
> Step 2 [Identification of the realizers of M] Find the properties (or mechanisms) in the reduction base that perform the causal task C.
>
> Step 3 [Developing an explanatory theory] Construct a theory that explains how the realizers of M perform task C. (Kim 2005, 101-102)

There are two other elements to Kim's theory that I would like to point out, which show the strengths of his theory. The first one is the principle of "local reduction", as Kim calls it. Naturally, someone's first reaction to his model might be to claim that the multiple realizability is a problem to Kim's model just as it was to the identity theory. To fend of this objection, Kim introduced the notion of "local reduction":

> Functional reduction, as I call it, can focus on the reduction of a mental property, or a group of them, for a specific population—that is, neural research on pain

[35] "For a functionally defined property M, any property in the base domain that fits the causal specification definitive of M (that is, a property that performs causal task C) is called a "realizer" of M." (Kim 2005: p. 101)

> will aim at *local* reductions, not a one-shot *global* reduction (as suggested by the Nagel bridge-law model). (2005: p. 25)

As opposed to the identity theory, we don't see the property of *being in pain* as a global, universal property, but instead opt for a more local approach, for example *human* pain, or *Martian* pain, etc. Tt can get even more local, fine grained and specific. I believe this to be a move that is completely in line with scientific practice, shows a more realistic picture of reduction, and avoids what is in my opinion a rather too demanding conditions of Nagel's "all-or-nothing", as Kim notes it (p. 102), condition for reduction.

The second element of Kim's theory concerns with the causal efficacy of the mental. One might object to this theory (or any reductionist theory in general) that it *reduces* causal powers of mental properties, something that we have encountered beforehand. Kim battles off this objection by positing *the causal inheritance principle*:

> According to the causal inheritance principle, the causal powers of an instance of a second-order property are identical with (or a subset of) the causal powers of the first -order realizer that is instantiated on that occasion. (1998: p. 116)

Sure, causal powers of mental properties might not be anything *over-and-above* the causal powers of their physical realizer, but this might not be such a bad consequence. Otherwise, we would have two options, either that of causal exclusion, or mere epiphenomenalism. This way, mental properties retain their causal powers, by inheriting them from their physical realizers.

There may, however, be some negative consequences to Kim's position. Even if we accept this kind of *local* reduction, what happens with the more general property of *being in pain*. There must be something bringing together Martian pain, human pain, dolphin pain, etc. and that something *is* the property of being in pain, somehow making a full circle back to type-type physicalism and the problem of multiple realizability. Kim has a way of avoiding this issue: instead of talking about pain as a general mental property, we should talk about it as a *concept*; merely a description or a designator of a property:

> (…) pains are pains because they conform to the definition of pain, not because they all share some hidden essence, like C-fiber stimulation or a pain quale. So there is the concept of pain, a concept given by its functional definition, but no property of pain, or being pain, that all pain instances have in common. (Kim 2008: p. 231)

However, Moore and Campbell (2010) argue that by doing so, Kim in fact becomes an *eliminativist*, this risking a possibility of losing mental causation:

> Conceptualized functional reduction leads to the elimination of mental properties in exchange for mental concepts alone. This also represents the abandonment of mental causation, for on this view mental properties do not exist, and something that does not exist cannot be causally efficacious. (p. 44)

I believe this is a somewhat wrong interpretation of Kim. First and foremost, just because Kim eliminates a more general, *universal* property of being a pain, that does not necessarily imply elimination of *pain* on all levels. Analogous to that, Kim argues that we might question the existence of property of being a table, but that by no means implies elimination of tables as such: "M as a concept stays, and individual instances falling under M are perfectly legitimate entities with causal-explanatory efficacy." (2008: p. 231) Furthermore, we still have local reductions, which allow for causal-explanatory efficacy, as Kim would have it. I believe this is the correct way to interpret Kim; instead of seeking one-to-one relation of a higher-level property of being in pain with its realizer, we should observe such pain merely as a concept (in my opinion a really useful concept). We might have ontologically eliminated general property of being in pain, but local pain*s* still exist, instantiated as human pain, Martian pain, perhaps even a robotic pain, and it has causal powers by simply being reducibly identified with some physical realizer.

6.3. What can be reduced? The problem of qualia

So far, I have not discussed this question, but I would like to touch upon it just briefly, before concluding this thesis. Generally speaking, the question is the following: what mental properties can be reduced? In other words, are there mental properties that *cannot* be reduced? To answer this question, I will stick to Kim and his view. In his opinion, the only mental properties that cannot be reduced are *qualia*, the qualitative aspects of mental states. The main problem here is that we intuitively want to believe that the colours we see, for example at the traffic light, *caused us* to stop or go; that the qualia do play causal role. At the same time, we also tend to believe there is something irreducible about qualia. So, how is that consistent with Kim's theory? Kim believes *qualia* are not reducible because they are simply not functionalizable. We can easily imagine a world with the inverted qualia[36] where people see

[36] Although the argument is originally suggested by Block, in this case I am using reiterations by Kim (2005).

red where we see green, and where they see green where we see red. Just briefly, his argument is the following:

> Such spectrum-inverted people would be as adept as we are in picking tomatoes out of mounds of lettuce and obeying traffic signals, and in general they would do just as well as we do with any other tasks requiring discrimination of red from green. If this is the case, colour qualia do not supervene on behaviour; two perceivers who behave identically with respect to input applied to their sensory receptors can have different sensory experiences. If that is true, qualia are not functionally definable; they are not task-oriented properties. (2005: p. 169-70)

In short, Kim believes that qualia are epiphenomenal, they are the mental residue; irreducible but also causally impotent. The jury is still out on what qualia actually are, how they relate to the physical world, and whether there is any basis for them in our brains. If you ask Kim on whether (and what) mental properties are reducible, he replies with the following statements:

> Yes and no: intentional/cognitive properties are reducible, but qualitative properties of consciousness, or "qualia," are not. In saving the causal efficacy of the former, we are saving cognition and agency. Moreover, we are not losing sensory experiences altogether: qualia similarities and differences can be saved. What we cannot save are their intrinsic qualities—the fact that yellow looks like this, that ammonia smells like that, and so on. (2005: p. 174)

Whether this is something a non-reductive physicalist can accept is debatable. But that is Kim's physicalism, or *something near enough*.

7. CONCLUSION

One fundamental question has been in the crux of this thesis: how is it possible for our mental states and properties to exert their causal influence in the world that is fundamentally physical? It has not been an easy task, but I have tried to provide a broad array of different options, with all their advantages and disadvantages. After seeing the motivation in terms of the importance of causal efficacy of our mental states, and the origin of this issue stemming way back from Descartes, the task was to see if we can save mental properties from being epiphenomenal. I have hopefully shown that epiphenomenalism is not an option, but that still did not clear up the idea of mental causation, so we had to opt for another option. The first set of options was subsumed under the idea of reduction, which saves causal efficacy of mental properties, but reduces them to physical properties. The first such option was the identity theory. The idea of this position was to save mental causation by identifying mental types with physical, thus avoiding the problem of physical closure and overdetermination. However, this position soon met its demise when Putnam argued for functionalism and multiple realizability. On the other side of the reductive spectrum, the idea of eliminative materialism was briefly depicted. This approach was also rejected, but it did prove itself to be rather convenient, thought-provoking and interesting. The second set of options was the non-reductive approach, depicted in terms of its doctrines and compared to Davidson's anomalous monism, which largely influenced the whole debate. However, Kim's argument of causal exclusion showed that non-reductive ambition to have their cake and eat it fell short. In turn, some non-reductive physicalists wanted to save mental causation without falling into the trap of causal exclusion by appealing to counterfactuals. At the end of that chapter, I argued that mental properties are *counterfactually relevant.* Nevertheless, for mental properties to be causally efficacious, they need to meet higher standards of causation, i.e. the generative or productive sense, something that counterfactual analysis cannot provide. Finally, Kim's own approach to solving the problem of causal exclusion was provided, the one I personally endorsed. His idea is one of functional reductionism, according to which mental properties are functionally defined and then (*locally*) reduced and identified with their physical realizers. Whether Kim is completely successful in saving mental efficacy without facing the problem of multiple realizability, without eliminating mental properties more generally is still open for debate. I believe this might be the best option we have so far.

Of course, there is still a lot to deal with. I also believe that there are many ways in which the debate on mental causation can turn up. One way in which this debate on mental causation can take its course would be in the context of the new forms of artificial intelligence. The progress in developing new artificial intelligence is a two-way street; the more we know about the functioning of artificial intelligence, the more we know about human behaviour and *vice versa*. Another route the debate on mental causation can develop would be hand in hand with neurosciences, as they still have a long way ahead of them. The scope of our knowledge about our brain and its functioning is still immensely low. However, as I said on different occasions, this should not stop us or impede us in making progress. As instead, it should be a motivation. I believe, if all cards are played right, philosophy could have a great role in the ages to come in terms of better understanding and explaining our nature.

BIBLIOGRAPHY

Berčić, B. (2012). *Filozofija*. Svezak drugi. Zagreb: Ibis grafika.

Berčić, B. (ed.) (2017). *Perspectives on the Self*. Rijeka: University of Rijeka.

Bregant, J. (2013). "The Problem of Causal Exclusion and Horgan's Causal Compatibilism". *Croatian Journal of Philosophy* 9 (3): 305-320.

Biondić, M. (2017). "Pučka psihologija: znanstvene perspective realizma, eliminativizma i instrumentalizma". *Filozofska istraživanja* 147(3): 559-578.

Burge, T. (1993). "Mind-Body Causation and Explanatory Practice". In: Heil and Mele. (eds.) (1993): 92-120.

Churchland, P. S. (1981). "Eliminative Materialism and the Propositional Attitudes". *Journal of Philosophy* 78: 67–90.

Davidson, D. (1970), "Mental Events". Reprinted in Davidson, D. (1980). *Essays on Actions and Events*. Oxford: Clarendon Press: 207-227.

Davidson, D. (1980). *Essays on Actions and Events*. Oxford: Clarendon Press.

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

Descartes, R., 1641/1993. *Meditations on First Philosophy*. trans. and ed. Donald A. Crass. 3rd ed. Indianapolis/Cambridge: Heckett Publishing Company.

Fodor, J. (1981). "The Mind-Body Problem". *Scientific American* 244: 114-125.

Fodor, J. (1989). "Making Mind Matter More." *Philosophical Topics* 17: 59-79. Reprinted in *A Theory of Content and other Essays.* Cambridge: MIT Press 1990: 137-160.

Garber, D. (2001). *Descartes Embodied*. Cambridge: Cambridge University Press.

Hall, N. (2004). Two concepts of causation. In J. Collins, N. Hall and L. A. Paul (eds.) *Causation and Counterfactuals*. Cambridge, MA: MIT Press: 225-276.

Hatfield, G. "René Descartes", *The Stanford Encyclopedia of Philosophy* (Summer 2018 Edition), Edward N. Zalta (ed.), https://plato.stanford.edu/archives/sum2018/entries/descartes/ (Last access: June 21st, 2019).

Heil, J. (2013). *Philosophy of Mind*. Third Edition. London: Routledge.

Heil, J. and Mele, A. (1993). (eds.). *Mental Causation*. Oxford: Clarendon Press.

Kim, J. (1984). "Concepts of Supervenience". *Philosophy and Phenomenological Research* 65: 153-176.

Kim, J. (1993). *Supervenience and Mind.* Cambridge: Cambridge University Press.

Kim, J. (1998). *Mind in a physical world*. Cambridge: MIT Press/Cambridge University Press.

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

Kim, J. (2007). "Causation and Mental Causation". In: McLaughlin, B. P. and Cohen, J. (eds.) (2007): 227-242.

Kim, J. (2008). "Reduction and Reductive Explanation: Is One Possible Without the Other?" Reprinted in Kim, J. (2010). *Essays in the Metaphysics of Min*d. Oxford: Oxford University Press: 207-233.

Kim, J. (2010). *Essays in the Metaphysics of Min*d. Oxford: Oxford University Press.

Kobašlić, M. (2016). Kontrafaktička teorija uzrokovanja. Završni rad. Sveučilište u Rijeci Filozofski fakultet. Odsjek za filozofiju, Rijeka. https://urn.nsk.hr/urn:nbn:hr:186:641895

Lepore, E. and Loewer, B. (1987) 'Mind Matters'. *Journal of Philosophy* 84: 630–42.

Lewis, D (1973). *Counterfactuals*. Oxford: Blackwell.

Libet, B. (1985) "Unconscious Cerebral Initiative and the Role of Conscious Will in Voluntary Action". *Behavioral and Brain Sciences* 8: 529–539.

Loewer, B. (2002). "Comments on Jaegwon Kim's Mind in a Physical World". *Philosophy and Phenomenological Research* 65: 655–662.

Loewer, B. (2007). "Mental Causation or Something Near Enough". In: McLaughlin, B. P. and Cohen, J. (eds.) (2008): 243-264.

Lycan, W. (2005) “A Particularly Compelling Refutation of Eliminative Materialism”. In: Johnson, D. M. and Erneling, C. E. (eds.) *The Mind as a Scientific Object: Between Brain and Culture*. Oxford: Oxford University Press, 197–205.

Malcolm, N. (1968). “The Conceivability of Mechanism”. *Philosophical Review* 77: 45–72.

McLaughlin, B. P. and Cohen, J. (eds.) (2007). *Contemporary Debates in Philosophy of Mind*. Malden, MA: Wiley-Blackwell.

Milojević, M. (2017). Embodied and Extended Self. In: Berčić, B. (ed.) *Perspectives on the Self*. Rijeka: University of Rijeka: 59-80.

Moore, D. and Campbell, N. (2010). “Functional Reduction and Mental Causation”. *Acta Analytica* 25: 435–46.

Patterson, S. (2005). “Epiphenomenalism and Occasionalism: Problems of Mental Causation, Old and New”. *History of Philosophy Quarterly*. 22: 239–57.

Putnam, H. (1967). “Psychological Predicates”. Reprinted as “The Nature of Mental States” in Putnam, H. (1975). *Mind, Language and Reality, Vol. II*. Cambridge University Press: 429–40.

Robb, D. and Heil, J. "Mental Causation", *The Stanford Encyclopedia of Philosophy* (Summer 2019 Edition), Edward N. Zalta (ed.), https://plato.stanford.edu/archives/sum2019/entries/mental-causation/ (Last access: June 21st, 2019).

Sesardić, N. (1984). *Fizikalizam.* Beograd: Istraživačko izdavački centar SSO Srbije.

Smart, J. J. C. (1959). “Sensations and Brain Processes”. *Philosophical Review* 68: 141–56.

Stoljar, D. (2010). *Physicalism.* New York: Routledge.