The Department of Philosophy

Theoretical Philosophy

THE INEVITABILITY OF REDUCTIONISM IN THE MIND-BODY PROBLEM

Finding convergence between functionalism and Type-Identity Physicalism

VALERY NKENGURUKE

C-UPPSATS

HT22

Supervisor: Andreas Stokke

Abstract

Type identity theory is a theory within physicalism that maintains that types of mental states or processes are identical to types of states or processes in the brain. This would imply mental types can't be encountered in physical structures other than the brain. But it seems to be the case that other biological creatures with different physical processes going on in their brain are able to experience a mental state such as pain. Even our brain, with its ability to undergo modification, functionally and structurally, in response to experience or injury, an ability called neuroplasticity, seems to have the ability to instantiate or realize the same mental state by two different neural events. This is the multiple realizability nature of the mental, that challenge type identity theory, but that functionalism seems to account for. Multiple realizability seems to provide a plausible basis for the irreducibility of mind to brain, and thus a good argument for why they should not be considered identical. I argue that when type identity theory claims that types of mental states or processes are identical to types of states or processes in the brain, this should be understood as claiming that types of mental states are identical to types of mechanisms in the brain, by which I mean types of interaction between fundamental particles that comprises the brain. I argue that two physical structures realizing the same mental state must be relevantly similar in the way their respective fundamental particles are interacting with one another. In other words, that similarity in function implies similarity in the way the physical realizers are organized. Therefore, the multiple realizability nature of mental states shouldn't render them irreducible or non-identical to specific types of interactions fundamental particles.

Keywords: Physicalism, identity, reduction, functionalism, multiple realizability.

Table of content

1. Introduction

There seems to be something ostensibly different between the mind and the brain[1], albeit phenomenological, otherwise there wouldn't exist a mind-body problem and their relation wouldn't warrant so many explanations as it has in the history of the philosophy of mind. Identity physicalism, being a version of physicalism, takes a monist position in this debate concerned with how to best understand the relation that obtain between mind and body (or brain as being part of the physical body), by asserting that the mind and brain are one (of one type of substance which is physical[2]) in that they are identical (they are one and the same). Its subtype Type identity theory holds that mental and brain states and processes, or events[3] can be categorized into types that can then be co-identified with one another. A type of mental state *is* nothing but a type of physical process or state that we call brain. It's this so called "nothing but" relation that characterize the identity and consequently a form of reduction in the theory.

This implies that if a type of mental state such as pain is nothing but, and thus identical to a physical process or event such as group of neurons firing (neural event) in the brain, it can't be encountered in places where that physical process, that is, where that specific group of neurons isn't firing. But it seems to be the case that it does. It seems to be the case that other biological creatures with different physical processes going on in their "brain" can experience a mental state such as pain. Even our brain, with its ability to modify itself, functionally and structurally,

[1] By brain I'm here referring to the entire nervous system but for simplicity I'm going from here on to only speak generally of the brain. One can also think of brain in places where it says body, for instance in "mind-body problem". When I speak of brain or body one can also think of them as instances of a physical entity. Sometimes I might thus speak of the relation between the mental and the physical as to note that I'm talking about the relation between the mental and the physical.

[2] There seems to be other ways of being a monist that doesn't necessarily claim reality to consist of one type of substance that is physical in nature. For example, one can ascribe to the dual-aspect monism, a thesis that take the mental and physical to be two different aspects of one substance or one can be a neutral monist by holding that the fundamental nature of reality to be neither mental nor physical, but rather neutral.

[3] I intend to use physical states and physical processes and evens interchangeably unless a distinction is warranted.

in response to experience or injury, an ability called neuroplasticity[4], seems to be able to instantiate or realize the same mental state by two different neural events. One can also speak of the future possibility of a computer brain construed entirely of different types of physical processes than the biological ones, but nonetheless is able to realize a mental state such as pain. It seems thus to not be the case that (at least all) types of mental states are identical or reducible to specific types of processes in the brain, such as types of groups of neurons firing. This multiple realizability nature of the mental seems to render mental states type identity theory implausible.

One of the traditional ways to account for this multiple realizability feature is to be a functionalist. Functionalism as an alternative to type identity theory argues that:

…what makes something a thought, desire, pain (or any other type of mental state) depends not on its internal [physical] constitution, but solely on its function, or the role it plays, in the cognitive system of which it is a part. More precisely, functionalist theories take the identity of a mental state to be determined by its causal relations to sensory stimulation, other mental states, and behavior. [5]

For a functionalist, a mind is that whichever can do the job of a mind analogous to that whichever can do the job of a table is a table regardless of the (physical) stuff it is made of. The job or role or function in functionalism is usually defined as an ability of any given object, may it be physical or not, to take in certain inputs, let's say excessive heat, and as a result output certain behavior such discomfort and removal of the object experiencing that heat. The object in this state would be said to be in pain, for example. It thus follows that this function or role can be fulfilled by different physical, but even non-physical states. Functionalism thus appears to be inherently non-reductionist (one could even say anti-reductionist). The multiple realizability nature of the mental, that challenge type identity theory but that functionalism seems to account for, seems thus to provide a credible basis for the irreducibility of the mind to the brain and thereby a good argument of why they aren't to be considered to be identical.

It's precisely this I want to question in this essay. I purport to question whether functionalism, at least in its physicalized form (since I will be operating under a physicalism worldview), actually manage to escape this reductionism that seems to be inherent in the type-identity theory as it purports to by accounting for the multiple realizability nature of the mind. I will do this by

[4] ADÃES, S. (2022, July 14). *WHAT IS NEUROPLASTICITY? —MECHANISMS OF FUNCTIONAL AND STRUCTURAL BRAIN PLASTICITY*. Retrieved January 24, 2023, from https://neurohacker.com/what-is-neuroplasticity-mechanisms-of-functional-and-structural-brain-plasticity

[5] Levin, Janet, "Functionalism", *The Stanford Encyclopedia of Philosophy* (Winter 2021 Edition), Edward N. Zalta (ed.), URL = https://plato.stanford.edu/archives/win2021/entries/functionalism/

discussing whether this multiple realizability thesis, this thesis that's commonly used to challenge the plausibility of type-identity theory, is enough to render mental states irreducible or non-identical to physical states. This will involve me arguing that functionalism in its physicalized form, instead of escaping reductionism by accounting for the multiple realizability of the mental as it purports to, seems to me to be converging with the type-identity theory in important and relevant ways. In a sense, I will thus be asking whether accounting for the multiple realizability of the mental allows functionalism to escape identifying or reducing[6] the mental to the physical. In this, I will be discussing whether the non-reducibility of the mental is best construed to be lying in its multiple realizability.

Here is how I'll go about this. Since type identity theory or type physicalism, as it is also known, is a physicalist theory, and hence it ascribes to a substance monism, I will in the first section provide some overview to how we should conceptualize the physical substance and therein how from a physical substance we get a complex array of entities in our physical reality. In light of that, I will explain what that conceptualization means in terms of how we group physical entities (brains, for instance) into types. Before I argue on how the identification of types of mental states to types of brain states needs to be done, I'll need to speak as well on the nature of the mind and how typing it happens. In this section, I will primarily make use of a paper from 2018: *Neural Correlates of Consciousness Meet the Theory of Identity* by Michal Polák and Tomáš. After this I will juxtapose this to the physicalized form of functionalism, and I will argue that it doesn't actually escape reductionism by accounting for the multiple realizability nature of the mind. Instead, it seems to be compatible in important ways to the reduction and identity relation that takes place in type identity theory. I will then in the concluding sections discuss some challenges that I see (type) physicalism and functionalism facing.

2. The physicalism in type-identity theory

Traditionally, one of the prominent ways of being a monist is to be a physicalist, which is a view that takes the physical to be the only substance of which the whole of reality is made. Being a physicalist means ascribing to the metaphysical thesis that asserts that everything is in their fundamental or substantial nature physical[7]. This claim that reality is monistic, and the

[6] I'm taking identity here to involve a reduction, that is to say that I regard identifying two things being the same as reducing one to the other or vice versa since one would be claiming that they're one and the same. More on this later.

[7] Stoljar, Daniel, "Physicalism", *The Stanford Encyclopedia of Philosophy* (Summer 2022 Edition), Edward N. Zalta (ed.), URL = https://plato.stanford.edu/archives/sum2022/entries/physicalism/

addition claim that the fundamental and substantial nature of this reality is physical, strikes me as being reductionistic because it reduces all real things to one kind, the physical kind. One must then ask: if everything is physical and if the mind is thereby physical, how is the physicality of the mental best to be construed, given the many ways things can be physical, as it is evident in the complex array of real things existing in the world.

Daniel Stoljar in his article *Physicalism* for the Stanford Encyclopedia of Philosophy[8] differentiate what he calls *the interpretation question*: "What does it mean to say that everything is physical?" from *the truth question*: "Is it true to say that everything is physical?" In this essay, I won't engage much the truth question, at least directly. What I say might indirectly have some relevant bearing on whether physicalism is true or not. Rather, this essay will mainly be associated to the interpretation question. Stoljar makes a further distinction. He divides the interpretation question into two sub-questions, which he terms the *condition question* and the *completeness question*. The condition question asks: "What does it mean to for something to be physical?" while the completeness question asks: "What relation or relations must obtain between everything and the physical if physicalism is to be true?" Defining this conditions that something must satisfy in order to count as *physical* in a non-circular manner, such that one can argue that everything including the mind satisfy that condition and thus solve the mind-body problem, is beyond my capacity and scope of this essay. I will in the next section speak more about how the concept of substance is to be understood, and thereafter I will discuss how I understand the physical substance fitting in that conceptualization. By doing so, I aim to clarify how the claim that everything is of physical substance should be viewed. But regarding the nature of this physical substance, we can for simplicity operate under the common conceptualization, at least according to the Miriam Webster online dictionary, that whatever the nature of the physical substance, it is that which relates to things perceptible through the bodily senses and subject to the laws of nature. In other words, things that are tangible or concrete or things characterized or produced by the forces and operations of the science of physics[9] (things having to do with matter or energy).

I will therefore mainly be concerned with the completeness question. As Stoljar reasons, "the completeness question holds fixed the issue of what it means for something to be physical and asks instead what relation or relations must obtain between everything and the physical if

[8] Stoljar, Daniel, "Physicalism", *The Stanford Encyclopedia of Philosophy* (Summer 2022 Edition), Edward N. Zalta (ed.), URL = https://plato.stanford.edu/archives/sum2022/entries/physicalism/
[9]Merriam-Webster. (n.d.). Physical. In Merriam-Webster.com dictionary. Retrieved January 3, 2023, URL=https://www.merriam-webster.com/dictionary/physical

physicalism is true." Thus, assuming we know what condition something must satisfy to be physical, what does it mean to say that everything satisfies that condition? I will therefore in this essay concentrate on how one can say that something is physical by saying that that something bears an identity relation to the physical, not by defining the condition that physical things satisfy[10]. The mind can thus on this reading be considered to be physical, in that it's identical to the physical brain.

2.1 The "physical" substance

In order to get us off the ground as to open for the possibility of understanding how everything, especially (types of) the mental, can appropriately be said to be (types of the) physical, I must briefly talk about how the physical substance of physicalism can be conceptualized so that it is possible for the mind to be reducible to, and thereby identifiable with it.

As Howard Robinson writes the article on Substance for the Stanford Encyclopedia of philosophy, of the "… six overlapping ideas that contribute to the philosophical concept of substance"[11] he mentions, I find three to be relevant for this essay. He says that substances are typified as:

i. being ontologically basic—substances are the things from which everything else is made or by which it is metaphysically sustained;

ii. being the paradigm subjects of predication and bearers of properties;

iii. being typified by those things we normally classify as objects, or kinds of objects;

This conceptualization is to be kept in mind as we proceed to the subsequent section. Because now we need ask what this means for our physical substance. How does (physical) reality end up looking the way it does, with its complex array of physical entities?

2.2 The layered nature of the physical.

[10] An important note to make here is that when I speak of relation or relation in physicalism, it's not to say that there exist two things that are relating to one another since physicalism, being a monist approach, holds that there exists only one type of substance. Rather I speak as such because, as I began, due the phenomenality of the mental, we tend to colloquially talk of them as separate entities. So, drawing from these conceptions in folk psychology, or in folk phenomenology or maybe even in folk ontology (if there is such thing) identity physicalism, as a philosophical theory, challenges them and says that they are not two distinct entities, rather they're one and the same thing.

[11] Robinson, Howard, "Substance", The Stanford Encyclopedia of Philosophy (Fall 2021 Edition), Edward N. Zalta (ed.), URL = https://plato.stanford.edu/archives/fall2021/entries/substance/

This section, as mentioned above, aims to explain how out of only physical substance you can end up with an array of complex and different things. According to the first defining feature of what a substance, a substance is “ontologically basic”, and from it everything else is made. If incorporated into the natural sciences, I understand "ontologically basic" to refer to the fundamental or elementary entities that the physical world seems to consist of. Since the atomic theory of matter, maybe even before (think of Democritus), the physical reality has been thought to consist of these infinitesimal entities commonly referred to as fundamental/elementary particles[12]. I think that it's these fundamental particles that fit best the description of what physical substance is because it’s from these that “everything else (all other bigger physical entities) is made. Going up the hierarchical levels[13] of the physical reality (so to speak) these fundamental particles coalesce into electrons, protons, and neutrons, which in turn coalesces into atoms, and then molecules and then chains of molecules, and then bigger structures such as biological cells or liquids or solid objects, and so on and so forth.

It should be noted then that any physical entity, unless it is one of the fundamental particles, that is to say, unless it is a substance and thus ontologically basic, is not one thing, rather it’s a composition of things, a composition of fundamental particles interacting in a certain way (see the iii criteria for substance). Take any common physical entity, tables, cars, human bodies, or water, according to the natural sciences, all those can be divided into smaller parts, which in turn are composed of even smaller physical things. This division, physics reports, goes on until we come to the fundamental particles, until we come to the ontological basic entities. This means that the physical realities we commonly conceive as things or objects are precisely more like many things interacting. Keeping up with this reasoning, when we speak of brain, we’re speaking of a composition of fundamental particles (ontological basic) interacting in a certain way as to bring about those characteristics that are normally associated with brains. It’s in this light that the physicality of the brain should be understood. Its physicality lies in that it is nothing more than a composition of ontological basic physical entities (fundamental or elementary particles) interacting. I bring this up because type identity theory is essentially about identifying types of mental states to types of states in the brain, which on this reading should

[12] As it is not clear according to our current understanding of physics whether it’s quantum particles or waves or energy that occupies that level, I will for the sake of simplicity use elementary or fundamental particles as a general designator to the physical stuff that exist at the bottom level of our layered physical reality.
[13] One can think of physical realities existing at hierarchical levels, where "entities" at any given level are entirely composed of entities existing at lower levels, except for the entities of the bottom level, which are ontologically basic (fundamental particles).

then be understood as identifying types of mental states to types of interactions between fundamental particles.

As these small fundamental particles interact with one another and thence coalesce together into larger structures, there emerges properties attributable only to those structure as wholes, not to their parts alone. For example, liquidity is a property of a liquid like water, that can't be found in its atomic parts, hydrogen, and oxygen. It's rather a property of their interaction. When the hydrogen and oxygen atoms interact with one another, causing each other to behave in a certain way, one of the results of their causal interaction is liquidity. This is why it is said to "emerge" at the level of composition of the water molecules. In a similar way, we can also view mentality as a property of the causal interactions between the parts of the structure within which it is encountered. In biological humans, it is thought to be the result of the causal interactions of structures such as (neuronal) cells in the brain. When the neuronal cells of the brain interact, their interactions tend to have characteristics we tend to call mental. I will argue that even though, let's say, liquidity is multiple realizable, its multiple realizability shouldn't, however, render it irreducible to the specific causal interaction that is common to all liquids. There are many types of liquids, but they share common physical attributes, precisely as liquids. The same goes for mentality. Despite its multiple realizability, I'm going to argue that it shouldn't be rendered irreducible and therefore identical to the causal interaction of these fundamental particles, a very complex causal interaction one might add.

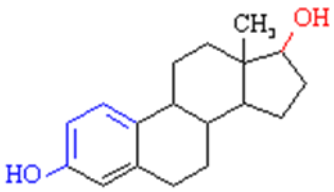
Before I talk about how mentality like liquidity can be thought to be identical to a specific type of physical interaction between fundamental particles, I want in the subsequent section to touch upon what makes two physical entities belong to the same type. What does it mean, for example, that all liquids belong to the same type? Given the fact that two liquids can be physically structured differently, what makes them identical in their "liquidity" as to make belong them to the same type? Analogously, given two different physical structures, what does it mean to say that they belong to the same type "brain"?

2.3 What makes two physical entities belong to the same type?

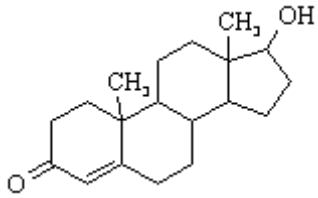
Following up on the earlier section, one can reason here that what defines any given physical object, a table for instance, it's not simply the number or kind of its parts, it's also the specific way those parts are causally interacting with one another, or put differently the specific pattern they are structured in. An example can be that two objects can be identical regarding the amount and types of fundamental particles they consist of but differ in the way those particles are

interacting or structurally organized. Thus, I take it that there are at least two ways that two physical objects can be identical or different. Their identity or difference can lie in the types and amount of the fundamental particles that comprises them, or in the way those particles are interacting. Most biological molecules are like this. Take for example testosterone and estradiol, the principal hormones which control the sexual characteristics of males and females, respectively. They seem to differ, among other things, in the number of group of atoms that comprises them, testosterone seems to have one -CH_3 more than estradiol and this amounts to a significant difference in their molecular properties.[14]

Estradiol (one of the female sex hormones):



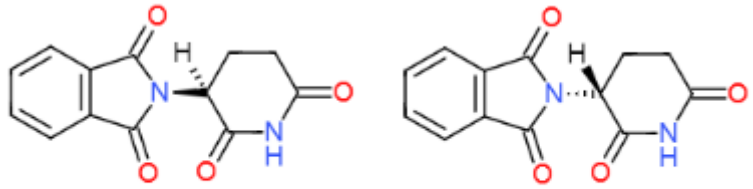
Testosterone (male sex hormone):



Thalidomide is an example of how difference between the ways the parts are interacting can lead to different properties between the whole structures.[15] There are two isomers, only differing in the way the atoms are situated in three-dimensional space, in their so-called stereochemistry. The (R)-isomer has a sedating effect while the (S)-isomer lead to horrifying birth defects in the children where the mother took the drug.

[14] StackExchange (2012, November 27). Are there notable organic compunds which look very similar but have very different properties? Chemistry. Retrieved January 26, 2023, from https://chemistry.stackexchange.com/questions/2633/are-there-notable-organic-compunds-which-look-very-similar-but-have-very-differe

[15] StackExchange (2012, November 27). Are there notable organic compunds which look very similar but have very different properties? Chemistry. Retrieved January 26, 2023, from https://chemistry.stackexchange.com/questions/2633/are-there-notable-organic-compunds-which-look-very-similar-but-have-very-differe



(S)-thalidomide (R)-thalidomide

What I'm trying to get at is the notion that for two physical compositions to be completely identical they'd have to be composed of the exact same number and kind of fundamental particles and those particles would have to be interacting with one another in an identical manner all the way up the hierarchy of physical reality, including not just the atomic level let's say, but also the molecular level, the cellular level (for biological entities), the organ level, and so on. But, if we stick to biology, two entities can be said to belong to the same type if there are enough *relevant* similarities in the number and types of the fundamental particles that comprise them[16] and, perhaps more importantly, if there is *relevant* similarity in the way those fundamental particles are interacting. So, belonging to the same type doesn't require being an exact copy, but rather to be relevantly similar, and exactly what makes two entities relevantly similar is what's difficult to pinpoint. The risk is that the typing of two physical structure into one type can be trivialized. The essential argument is going to be that if they're functioning in similar ways, then they must share relevant physical similarities that allow them to function in similar fashion. Think of how despite the difference between them, testosterone and estradiol are still considered to be of the same type, that is sexual hormones[17]. And despite these hormones differing in quantity between females and males, nevertheless a certain female and a certain male can be considered to belong to the same type, both can be human beings for example[18]. Think also of liquids. Liquid is one of the four primary states of matter, which is to say that it's a specific type of interactions between fundamental particles. "Being a liquid" is thus being in a specific type of physical state, akin to how "being a mental state" amounts to being in a specific *type* of physical state.

[16] Since, at least for biological beings (possibly for all physical beings), the fundamental particles tend be the same, what tend to distinguish two beings apart is the way those particles are interacting with one another.

[17] And I don't mean that they don't belong to the same type because they have the same function in females and males, respectively, rather that have the same function because they're physically structure in identical manners. More on this later.

[18] It should also be noted that, from this follows that, if physicalism is true, there can't be any difference in between two structure wholes, female, and male for example, that isn't reflected in the difference between the way the (fundamental) parts that comprises them are interacting with one another (estradiol and testosterone for example).

Why is this relevant? It's relevant regarding how the typing of the physical is made. Two physical entities can be said to belong to the same type (or not) depending on how the typing is done. Connecting this to the previous section, the reasoning above is also relevant because one can argue, which I am and will do so more in later sections, that the fact that the biological type of "human" is multiple realizable in both females and males shouldn't render it irreducible and therefore non-identical to a specific type of interactions between fundamental particles that comprises them. Analogous to how "being a liquid" doesn't stop being a specific *type* of physical state just because it is multiple realizable. It can be argued that there are sufficient relevant *physical*[19] similarities between how the fundamental parts that make up a human female (or a certain liquid) interact with each other and how the fundamental parts of a human male (a different liquid) interact with each other so that the type "human" (the type "liquid") can be reduced to and identified with that type of interaction that must be common between them. In the next section I want to speak on how the typing of the mental and the brain in type identity theory, when done appropriately, leads to the identity between types of mental states and types of brain states. Before I do that, let me, as I just did for the physical, speak briefly on the nature of the mental.

3. The nature of the mental.

Similar to how "physical" doesn't make itself easily defined, what mind is, is also hard to define. But similar to how defining what "being physical" means is not crucial, since in our case it's the relation to the physical that's of greater relevance, not getting the exact definition of "being mental" shouldn't impede much on my reasoning. The most common general classes seem to sensory states s like visual or auditory sensations), propositional states like beliefs and desires, intentional states, and conscious states. Whatever the many conceptualizations of mind may be, what seems to be common to all is their phenomenality or their subjectivity. But this I'm mostly referring to the fact that the subject of mental states (the person experiencing them), has privileged epistemic access while others can only infer their existence and nature from outward signs.

4. The nature of the identity in type-identity theory

[19] It has to be again noted here that it's physical similarity, not simply functional similarity. Or that functional similarity implies physical similarity.

By diving into this paper *Neural Correlates of Consciousness Meet the Theory of Identity* by Michal Polák and Tomáš Marvan[20] I want to first lay out a plausible way of understanding how mental and brain states are grouped into types. And then by discussing what it means for types of mental states to be identical to states or processes of the brain I mean to flush out the claim that I've been making, namely that the multiple realizability nature of mental states shouldn't render them irreducible or non-identical to brain processes. By doing so, I intend to show that type identity theory converges or at least is compatible with functionalism, in its physicalist form.

4.1 The typing of the mental

How is the typology of mental states and neural states, respectively, to be done so that, appropriately, an identity relation between them can be established? Given the phenomenal or subjective nature of mental states, Polák and Marvan reason that mental types "…are to be typed by their phenomenal features alone. They, too, confirm that this is a subjective endeavor due to the private nature of mental states. They write, "The criteria for distinguishing phenomenal types are subjective, because these mental states are not publicly available."[21] But nonetheless, there seems to be commonly shared characteristics that makes the typing of mental types possible. Think again of the common types mentioned above that mental states tend to divide themselves in. Sensory sensation is for instance one of the general types, in which we find visual, auditory, or tactile sensations. Conscious states can also be seen as one of those general categories in which we can find all types of subtypes.

The authors then propose "a hierarchical classification of phenomenal and neurophysiological types, spanning multiple levels of varying degrees of generality". At the top of the hierarchy, you find the most general types, at the bottom the most "minimal" types and then there is the "in-between-types". The most general types corresponding to above-mentioned general types. The minimal types are defined as those in which a subject can't "distinguish any phenomenal difference between at least two different subjective experiences". The picture below shows how the typing of sensory states might be done.

[20] Polák, Michal & Marvan, Tomáš. (2018). Neural Correlates of Consciousness Meet the Theory of Identity. Frontiers in Psychology. 9. 1–13. 10.3389/fpsyg.2018.01269

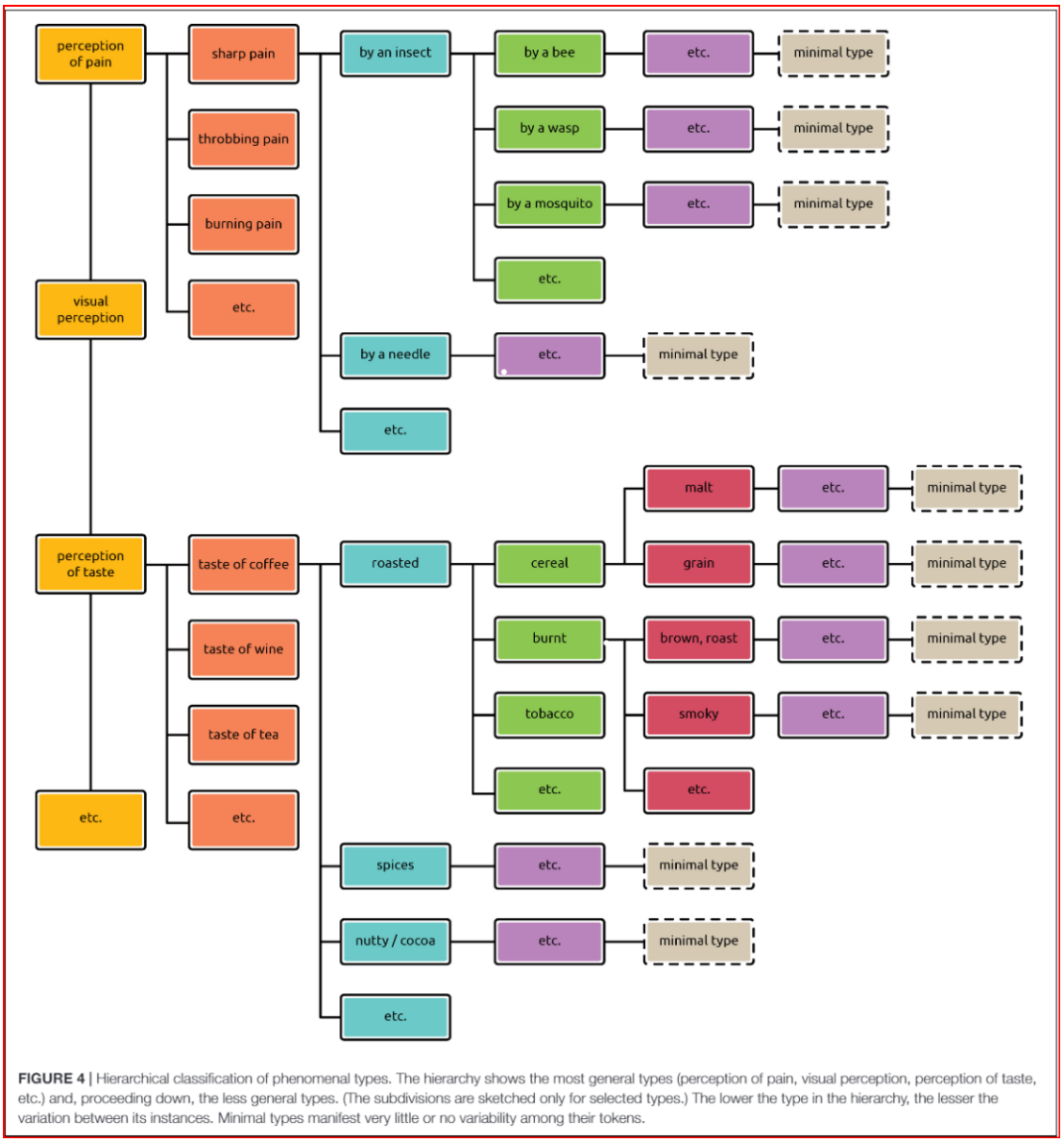


FIGURE 4 | Hierarchical classification of phenomenal types. The hierarchy shows the most general types (perception of pain, visual perception, perception of taste, etc.) and, proceeding down, the less general types. (The subdivisions are sketched only for selected types.) The lower the type in the hierarchy, the lesser the variation between its instances. Minimal types manifest very little or no variability among their tokens.

4.2 The typing of the brain

The typing of neural (brain) types, as opposed to the typing of phenomenal mental ones, they argue, happens through different means. Often this involves locating the activated brain structures and neural populations correlating with phenomenal tokens in the same type, through repeated measurements. Take multiple tokens of the type "pain perception", for example. As a result of scientifically studying mental states such as pain neuroscientists do discover through multiple measurements brain structures that regularly gets activated during those perceptions, and they do also find specific neural structures that tend to correlate with those perceptions reliably and regularly, hence the term Neural correlates of consciousness or NCC. Neural correlates of consciousness (or NCC) refer to the relationships between mental states and neural (physical) states and constitute the minimal set of neuronal events or process sufficient for a

specific conscious percept"[22]. Their technical definition is that an NCC of a phenomenal type *P* or mental type) is:

a type of neural event or process N such that there is a mapping, where (i) each neural token n_i of N is minimally sufficient for a phenomenal token p_i of P, and (ii) where all and only neural tokens of N instantiate a neural mechanism M, such that M is a sufficient condition for being an NCC of P.

They explain that these "regular mind-brain correlations form an evidentiary basis…" that show that certain states and processes of the mind (consciousness) correlate regularly and reliably with certain states and processes of the brain but that these NCCs "… can be used in support of different metaphysical accounts". I understand this to mean that what the empirical data can conclusively prove is only correlation, and there needs to be incorporated a theoretical presumably philosophical framework, as they themselves reason, to better make sense of the results in neuroscience and thus elucidate the ontological "nature of the relationship between subjective mental/phenomenal states and neurophysiological (physical) processes". As Place argued years ago that the identity between consciousness and brain can't be denied on logical grounds, rather it's something for the empirical sciences to establish[23]. Polák and Marvan thus appeal to the same theory, referring to it as an "elegant tool for interpreting the results of neuroscientists' lab-based mind-brain correlation measurements.

One of the ways these correlations have been theoretically understood, according to these authors, is that these NCCs *causes* the states and processes of consciousness that they supposedly correlate with. They refer to multiple papers where NCCs are explained in causal terms. Intuitively as it may seem, to "go from correlation to causation" as it is a typical move of the sciences, they argue that this is far less plausible than reasoning that these mind-brain correlations point to identity relations. This is how they reason: If a neurophysiological state or process causes a phenomenal/mental state of consciousness, then it must antecede it, and it must be different from that state, because causes and effects are always distinct. Not only that, since we're working under the assumption that physicalism is true, then it needs to be the case that every phenomenal/mental state is physical, that is, it must be instantiated or realized physically, i.e., by some brain (physical) state(s) or process. So, if it is thought that the NCC cause the mental state, and if it is the case, according to physicalism, that the mental state must be instantiated or realized by a physical state or process distinct from its (neural) cause, it follows that for there to exist one mental or process you'd need two physical states or processes, one to

[22] Polák, M & Marvan, T. (2018). *Neural Correlates of Consciousness Meet the Theory of Identity*. Frontiers in Psychology. 9. 1–13. 10.3389/fpsyg.2018.01269
[23] PLACE, U. T. (1956), *Is Consciousness a Brain Process*? British Journal of Psychology, 47: 44–50

cause it and the other to instantiate it. The first is there to cause the second, which in turn is there to anchor or realize it in the physical world (see picture below).

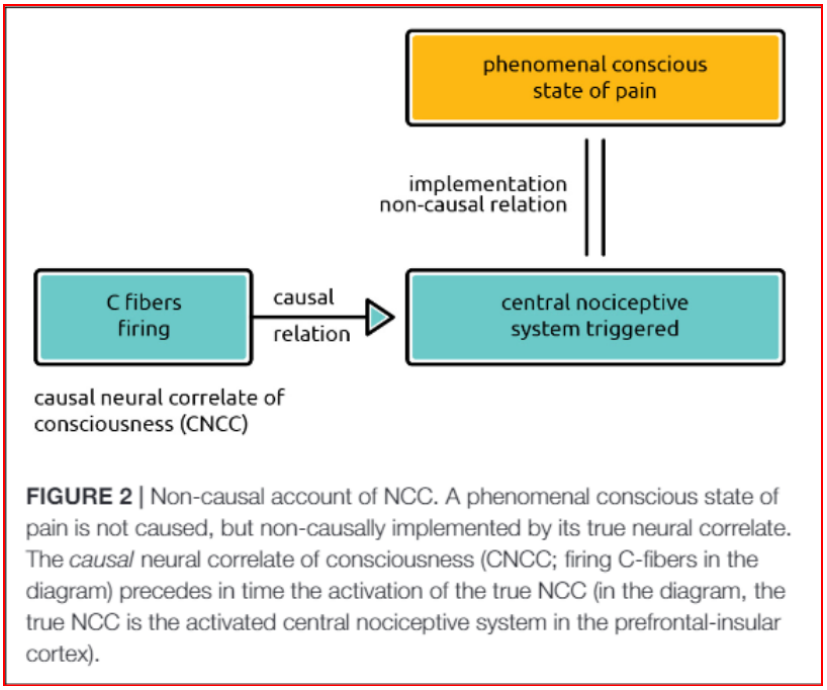


FIGURE 2 | Non-causal account of NCC. A phenomenal conscious state of pain is not caused, but non-causally implemented by its true neural correlate. The *causal* neural correlate of consciousness (CNCC; firing C-fibers in the diagram) precedes in time the activation of the true NCC (in the diagram, the true NCC is the activated central nociceptive system in the prefrontal-insular cortex).

The authors explain that the second neural process is the one to be considered the real NCC. Even though the first one would also regularly correlate with the state of consciousness in question, it is instead referred to as the Causal Neural Correlate of Consciousness (CNCC). They mean that neuroscientists in their empirical study of consciousness are primarily interested in the *realizing* NCCs, even though the causal ones are relevant. They give the example of C-fibers commonly brought up by philosophers of mind in these questions. They explain that "these firings are, in fact, the CNCCs of some kinds of pain sensations, because they are the causal antecedents of these sensations", they come before the actual *sensation* of pain. Another example they give is the processes that occur in the retina and in the optic nerve, being CNCC of visual perceptions. Those processes antecede the actual visual perception, they're part of the causal chain, but they're not the actual NCC, that is, the actual neural composition that *is* the visual perception.

In the figure below, the authors go on reasoning that one can talk about horizontal causation, as a causal relation that obtain between the true NCC (the true neural substrate of the mental state) and its causal antecedents and consequents (the CNCCs), as opposed to vertical "causation", the relation that obtains between the NCC and its correlate phenomenal/mental state. I'm putting the vertical "causation" in quotes to highlight the previous point, that it's not an actual causation. Rather, it's an identity relation. Whatever is going in N_2 is the actual mental state. N_2 doesn't cause the mental state, it *is* the mental state. They highlight the important notion that

the neural event N_2 can still be a complex neural process constituted by multiple causal electrophysiological and chemical processes. But one should not confuse those interactions that may be going on in N_2 and horizontal causal interactions between different Ns (different neural events) and the vertical "causal" relation. It's only the vertical "causal" relation between the whole neural system N_2 and its correlate mental state that is equivalent to the identity relation. The claim being made here is that any given mental state will be identical to its NCC. In this subsequent section, I want to flush out how this identity should be understood.

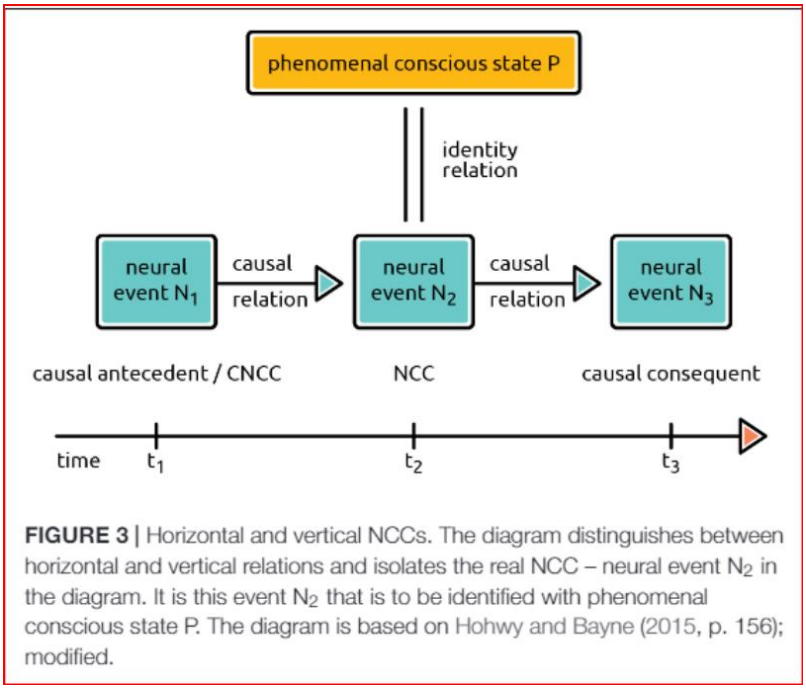


FIGURE 3 | Horizontal and vertical NCCs. The diagram distinguishes between horizontal and vertical relations and isolates the real NCC – neural event N_2 in the diagram. It is this event N_2 that is to be identified with phenomenal conscious state P. The diagram is based on Hohwy and Bayne (2015, p. 156); modified.

4.3 Identifying mental types to brain types

Although, as the authors argue, the neuroscientist "expect a degree of systematicity in how the NCCs of various types of phenomenal states are localized", it's not precisely the specific active brain structure or the specific neural population where the NCC is localized during a given type of phenomenal experience that can be said to be identical to the mental type. Consider again the given definition of the NCC, being "the minimal set of neuronal events or process sufficient for a specific conscious percept". The emphasis here is on "neuronal events or process". Polák and Marvan argue that scientists aren't primarily interested in the specific neural structures, but rather the way those structures are structured. I understand this to mean the way the fundamental parts in them are functioning. They talk about that due to the phenomenon such neural plasticity,

which is the ability of neural populations in the brain to adapt and change through growth and reorganization, the localization of neural types can't be exhaustive or absolute. They reason that what scientist aren't interested when looking for NCCs, it's not precisely any specific neural structure that's of interest, rather, they say, it's "the neurophysiological process, the mechanism, whatever it may be, sufficient for having [implementing or realizing] a phenomenal state". They mean that the purpose of localizing the active brain structures during a phenomenal/mental experience is to "differentiate brain areas due to their functional properties so that it is subsequently possible to concentrate on the neural pattern-properties themselves".

I interpret this to mean that it's the way those neurons are interacting with one another that's of interest, not those exact group of neurons themselves. This means that the identity in type identity theory lies in the fact that any given mental type *is* a type of interaction between neurons, and since those neurons themselves are specific types of fundamental particles interactions, a mental type, like pain for instance will thus be identical to a specific type of interaction between fundamental particles. Polák and Marvan conclude thus that being in certain a mental state, for *appropriate mechanisms*, consists in, or just is, being in a certain physical state (remember the liquid analogy, a certain liquid is nothing more than a physical state). Accordingly, whichever physical structure, this type of mechanism or interaction is encountered, that structure will be in that mental state. Therefore, the multiple realizability of any mental state shouldn't render it irreducible or non-identical to a certain type of interaction of fundamental particles because the mental state in question would be describing that type of interaction, that type of physical state.

5. On identity theory converging with functionalism

As stated, functionalism states that mental states are functional kinds, indicating that what makes a given structure, be it physical or non-physical, to be in a mental state, depends on whether it is functioning in a defined "mental" way. From this following that you must functionally define the mental state first to know if any structure is functioning that way. For the functionalist this involves holding that any object which, upon certain inputs, outputs certain mental-like behavior, is in a mental state. Analogous to how any object which, upon certain inputs, like something being put on it, is able to output certain behavior, like supporting the said thing, that object is a table, that is if "being a table" is defined as the ability to support things put on top of you. Thit is to say that whichever object can support something put on it is a table. It's conceivable that many objects will be able to function this way, indicating that the function

of "being a table" is multiple realizable. If this is applied to physicalism, then the only type of objects that can perform the function "being a table" are physical. "Being a table" remains multiple realizable, but it gets its multiple realization gets limited to the physical type.

If to be a table means, given certain input, to output certain behaviors, what would it mean for any physical entity A to be a table? Call to mind how we conceptualized physical entities, a physical entity is simply a composition of fundamental particles interacting in certain way. For A to be table would thus mean, given a certain input, for the fundamental parts that comprises it to output certain specific behavior, that is, to interact with each other and with the fundamental particles of the object put on top of it, in a specific way. These types of fundamental particles interactions will vary between tables, but they will be enough relevant similarities among the tables such that they can be grouped into types. Consequently, wherever we have that *type* of fundamental particles interactions (that type of pattern of organization of the fundamental particles), we have a table. Hence it can be said that "being a table" is identical to a specific type of fundamental particles interaction.

This is the point I was alluding to earlier. I'm arguing that similarity in physical function between two physical entities must imply similarity in the way their respective fundamental parts are interacting with one another. Regarding mental states, I'm arguing then that, is that even though mental states are realizable in multiple physical structures, these structures must be relevantly similar in the way their fundamental parts are interacting if they are to realize mentality. If the physical structures that are realizing the mental states are too dissimilar, the mental states will also be dissimilar, maybe even to a point where given enough dissimilarity between their realizers, the mental states can't be said to be of the same type. Consider the mental state of pain in biological brains and future robots. If the NCCs of the mental state between the biological brain and the robot's brain aren't similar enough, that is to say, if the fundamental parts of their brain responsible for realizing pain aren't interacting in relevant similar ways, the pain they experience will differ enough to the point where they shouldn't be said to belong to the same type[24]. This is also because by defining pain functionally, you are already defining it as specific type of interaction, as argued above, as an ability to, given certain inputs, output certain behavior, implying that whenever this type of interaction is exhibited you have pain.

[24] Remember that difference in the ways two physical entities are structured (testosterone and estradiol examples), leads to difference in the way they function. Implying that too much difference entails too much dissimilarity in function.

A different example of this can be how different multiple music instruments can create the same type of note (functional kind), by making air molecules vibrate at identical frequencies while at the same time sounding different. The musical note will be identical to the common frequency, i.e., the type of interactions between air molecules that's common to all. Or analogous to how all molecular structure having the property of solidity need to resemble one another in how their microstructures causally interact so that they are densely packed and kept together. Solidity like mentality would be thus identified to a type of causal interactions between fundamental particles. From this follows that, mentality like any other functional kind warrants a specific type of causal interaction between the microstructure of the object that is realizing it.

6. Conclusion

My aim in this essay was to try to bridge the gap between type-identity theory and functionalism. The multiple realizability nature of mental states commonly issued at type-identity theory as a stumbling block can be resolved by conceptualizing the physical type not as a type of physical object qua object, rather as a type of interactions between particles that comprises the physical "object". Thus, since what differentiate physical objects is mainly how the fundamental particles are structurally and causally interacting, it makes sense therefore to type objects depending on the type of interactions they share. Saying that a type of mental state is identical to a type of physical structure, say brain, would thus simply mean that a type of mental state is identical and reducible to a specific type of interactions between the fundamental particles that comprises the physical structure. And by identity, I mean that the type of mental state *is* a type of fundamental particles interactions. Stated differently, mental states are nothing but, or nothing more than (herein lies the reduction), a specific type of physical process. When a (physicalist) functionalist defines mental states as functional kinds, by doing so they're defining them as a way of behaving physically given certain inputs. I argued that two physical structures functioning similarly, can't be doing so arbitrary, rather they must also be physically behaving similarly. In other words, that similarity in function implies similarity in the way the fundamental particles of the objects are interacting. Consequently, any two physical objects, whether they be of wood or metal, or any other physical composition will be similar in the way their fundamental parts are behaving if they're both functioning as tables. The same goes for brains, whether they be biological or machine, if they're both realizing the same type of mental state, there must be enough relevant similarity in how their fundamental particles are behaving that legitimize the positing of an identity relation between that type of mental state and the state(s) or process(es) in said brains.

7. Discussion

A case can be made that I've misconstrued functionalism by revising it in physical terms, since the functionalist definition of the mental, as functional kinds or roles, renders mental states abstract or at least intangible. This is reflected in the fact that these types are multiple realizable, that they can be encountered in different physical structures. By arguing that functionalism converges with type identity theory or type physicalism in that for type identity theory mental states are types of physical interactions between fundamental particles, I haven't argued away multiple realizability and the problem it brings. The problem is that mental states are rendered abstract, or at least not entirely physical because one of the characteristics of the physical, according to the online Merriam-Webster dictionary, was that the physical is concrete or tangible. But concrete things aren't multiple realizable. An additional problem would be that mental state, which were said to be subjective, and thus private, can be encountered in other physical structures. But it seems to be the case that my pain, my headache let's say can't be multiply realized.

One can argue against this by saying that it's the tokens of the mental states that are in fact identical to tokens of physical states. We only generalize them into types because it's characteristic of the sciences to systematically generalize systematically token into types, particulars into generals. My specific pain, my headache let's say is a concrete interaction of particles. It's just that I'm able to think that specific interaction in isolation of the given particles that's realizing it. Just because I can abstract "headache" from its specific physical interactions and type it as type of a mental state "headache" of the general type "pain", doesn't mean that it stops being a concrete or tangible interaction between a group of fundamental particles. But also, regarding its private nature, one can reason that this type of pain, namely a headache, can be encountered in other physical structure, just that it wouldn't be me, which is to say that it wouldn't be this particular physical structure, realizing it and thus experiencing it.

But, and here is the crux of it, can an argument be made that there is something more to the physical process that is the headache? As mentioned, a physical entity is not just a collection of fundamental particles, it's a collection of fundamental particles organized or patterned in a certain way. It can thus be argued that any physical entity is a combination of both the particles that comprises it and the specific pattern in which those particles are interacting. Since the fundamental particles seems to be more or less be of the same type across different objects, then when we perceive two objects to be different, we are perceiving the difference in the way their particles are patterned. In other words, part of me seeing a table, is "seeing" the pattern that the

fundamental particles are arranged in, maybe more so than the particles themselves. For Aristotle this would be the distinction between matter and form[25] (maybe even the distinction between substance and its properties or maybe even the distinction between concrete and abstracts). Just as the (physical) substance can't help but have properties, so can't matter (fundamental particles), help but come packaged in intelligible/knowable patterns. We tend to thus see matter patterned as tables, books, brains, hearts, atoms, or molecules. The question then becomes, if to be physical is to be nothing more than the interaction of fundamental particles and if these interactions can help but be patterned in a certain way, and if it's true that it's these patterns that we recognize when we "see" physical entities, in what sense are these patterns physical? If I hear a certain music note for instance, let's say C, I can argue that this musical note is nothing but air molecules vibrating with a certain frequency and in that sense the note is physical. But is the frequency, being a mathematical property of the vibration of those air molecules, also physical?

This is to question the whole worldview of physicalism that claims that the whole reality is physical. It might be true that the physical substance is physical, but can that also be said of its properties? In Aristotle terms. if reality is matter and form, and if the nature of that matter is physical, what's the nature of the form? On one hand I think that type physicalism must grant some form of irreducibility of the mind to the brain because if the physical is known through its form, and if it can be argued that the form is irreducible to the physical matter, then the mind can't entirely be reduced to the brain, that is if the mind can be seen as a form of the brain. But this, on the other hand, might risk making the mind too abstract, something I see the functionalist who claim the mind to not be reducible to the brain do. This is why I think that if the functionalist must in turn grant that my headache *is nothing more than* a specific process going on in my brain. This shouldn't take away the fact that this type of pain can't be encountered in other physical structures, but as I have argued if they're to belong to the same type of mental state "headache", the physical process that realizes said mental state must be share relevant similarities.

With an addition example, I wish to problematize (type) physicalism further. Think of the reality of truth (or even intelligibility of rationality or logic). When the mind grasps a truth, the type physicalist would want to say that there is a specific type of physical process taking place that *is* the "the grasping of the truth". That is to say, that the mental processes of my

[25] Ainsworth, Thomas, "Form vs. Matter", *The Stanford Encyclopedia of Philosophy* (Summer 2020 Edition), Edward N. Zalta (ed.), URL = https://plato.stanford.edu/archives/sum2020/entries/form-matter/

understanding that a certain proposition is true is nothing more than a certain interaction between fundamental particles in my brain. The propositional mental state of holding a certain belief gets to be physical. But what about the truth or falsehood property of the said propositions in said mental state, are they also physical? If truth is considered as a relation between two physical things, two propositions let's say (that is, if one regards propositions as physical entities[26]), does that make truth also physical? What about rationality or intelligibility? Even though it can be argued that wherever we find truth, rationality, or intelligibility, we will find specific types of physical interactions or relations between the fundamental particles, it seems inapt to hold that the properties of those interactions, being true, rational, or intelligible, are also physical.

As a substance-monistic worldview, physicalism carries with it the risk that the only things that exist absolutely are the fundamental particles. Everything else is different modalities of those particles. But for something like that to even be asserted something like truth or intelligibility must be real, independent of any specific instantiation. This is the issue with these types of relations. Another example can be, the graspability of a mug, which is neither in the mug nor in me who is doing the grasping, rather it's a relation between me and the mug, that seems to be real, independent of any specific physical relata like me and the cup. In what sense is the graspability physical? These relational entities, such as truth, intelligibility or graspability, seems to differ from the mind in that they seem to be entirely abstract, i.e., they seem to exist entirely in the world of "forms" whereas the mind, in its being identical to the brain as I've argued, seems to be a combination of matter and form. The mind seems to participate in both. Thus, I can concede that there is an aspect of the mind that gets disregarded or mischaracterized by asserting that it is (entirely) physical. However, this would also apply to the brain or any other physical entity that's not ontological basic. Essentially thus claiming that the only entities that get to be entirely physical are the fundamental particles (maybe not even those). Because as soon they get together in a specific pattern, as soon they coalesce and form (note the verb "form" here) atoms, for example, then the atom doesn't get to just be physical anymore. This is because an atom isn't just a group of fundamental particles, it is an intelligible (knowable) patterned group of particles (it's matter with a specific form, a form that can be discovered and studied) so it's both particles plus the intelligible form they are patterned in. This will also be

[26] This is in itself problematic because even though it can be argued that propositions are always physically instantiated, by sound in speech, by ink on a paper in writings and alike, in what sense are they physical?

true, maybe even more so, of the brain. Mental states would, on this reading, be just matter with specific types of forms.

8. References

- ADÃES, S. (2022, July 14). *WHAT IS NEUROPLASTICITY?—MECHANISMS OF FUNCTIONAL AND STRUCTURAL BRAIN PLASTICITY*. Retrieved January 24, 2023, from https://neurohacker.com/what-is-neuroplasticity-mechanisms-of-functional-and-structural-brain-plasticity
- Ainsworth, Thomas, "Form vs. Matter", The Stanford Encyclopedia of Philosophy (Summer 2020 Edition), Edward N. Zalta (ed.), URL = https://plato.stanford.edu/archives/sum2020/entries/form-matter/
- Levin, Janet, "Functionalism", The Stanford Encyclopedia of Philosophy (Winter 2021 Edition), Edward N. Zalta (ed.), URL= https://plato.stanford.edu/cgi-bin/encyclopedia/archinfo.cgi?entry=functionalism
- Merriam-Webster. (n.d.). Physical. In Merriam-Webster.com dictionary. Retrieved January 3, 2023, URL=https://www.merriam-webster.com/dictionary/physical
- PLACE, U. T. (1956), *Is Consciousness a Brain Process?* British Journal of Psychology, 47: 44–50
- Polák, Michal & Marvan, Tomáš. (2018). *Neural Correlates of Consciousness Meet the Theory of Identity*. Frontiers in Psychology. 9. 1–13. 10.3389/fpsyg.2018.01269
- Robinson, Howard, "Substance", The Stanford Encyclopedia of Philosophy (Fall 2021 Edition), Edward N. Zalta (ed.), URL = https://plato.stanford.edu/archives/fall2021/entries/substance/
- Smart, J. J. C., "The Mind/Brain Identity Theory", *The Stanford Encyclopedia of Philosophy* (Winter 2022 Edition), Edward N. Zalta & Uri Nodelman (eds.), forthcoming URL = https://plato.stanford.edu/archives/win2022/entries/mind-identity/
- StackExchange (2012, November 27). Are there notable organic compunds which look very similar but have very different properties? Chemistry. Retrieved January 26, 2023, from https://chemistry.stackexchange.com/questions/2633/are-there-notable-organic-compunds-which-look-very-similar-but-have-very-differe
- Stoljar, Daniel, "Physicalism", The Stanford Encyclopedia of Philosophy (Summer 2022 Edition), Edward N. Zalta (ed.), URL= https://plato.stanford.edu/cgi-bin/encyclopedia/archinfo.cgi?entry=physicalism