Hard problems in the philosophy of mind

Alexandros Syrakos[1]

1 University of Cyprus

Funding: No specific funding was received for this work.

Potential competing interests: No potential competing interests to declare.

Abstract

The mind is our most intimate and familiar element of reality, yet also the most mysterious. Various schools of thought propose interpretations of the mind that are consistent with their worldview, all of which face some problems. Some of these problems can be characterised as "hard", not in the sense of being difficult to solve (most problems concerning the mind are difficult), but in the sense of being most likely insurmountable: they bring to the surface logical inconsistencies between the reality of the mind as we perceive it and the fundamental metaphysical tenets of that particular worldview, thus putting the latter in danger of being disproven. This essay focuses mainly on the hard problems that the author considers to be of greatest importance for physicalism, the currently prevalent worldview. Nevertheless, some of these hard problems pertain also to other views such as panpsychism. In the author's opinion, the hardest and most profound of these, pertaining equally to physicalism and to panpsychism, is the one discussed in Section 4: the particular subjective first-person viewpoint that defines a particular person can be found nowhere in the universe except in that person itself; all outside entities (physical or mental) are equally neutral towards the "particularity" of that person, which therefore cannot be explained as arising from any combination of such outside elements. Therefore, a person is a simple substance. Other hard problems discussed concern the physical explanation of conscious experiences and the physical explanation of meaning, while their repercussions with respect to free will and ethics are also examined. Although these latter hard problems have already been discussed elsewhere, a somewhat fresh perspective is given here by someone who is not a professional philosopher but a physical scientist.

Alexandros Syrakos*

Department of Mechanical and Manufacturing Engineering, University of Cyprus, P.O. Box 20537, 1678 Nicosia, Cyprus

*alexandros.syrakos@gmail.com, alexandros.syrakos@gmail.com

1. Introduction

Physicalism, the worldview that reality is physical at its most fundamental level and all its aspects can be ultimately explained by the fundamental laws of physics, is arguably the mainstream view in modern developed societies, particularly

among people of higher education. Although forms of physicalism or materialism have existed since antiquity (e.g. the views of Democritus [1]), its modern prevalence is in large part due to the impressive progress of science during the recent centuries, manifested in terms of both explanatory and predictive capability, and evidenced by spectacular technological advances.

Another factor that may have contributed to the prevalence of physicalism is the abuses associated with previously prevalent worldviews, which acknowledge purported immaterial aspects of reality such as God and souls. So, contemporary zeal towards physicalism may, at least in part, stem from the reasoning that "the enemy of my enemy is my friend": if physicalism is true, then the religious worldviews are false.

However, although this "friend" helps to reject the abuses and control imposed by those worldviews and provides some sort of freedom, the price to pay is very high: one has to accept that the intuitive impression we have of ourselves as individuals (simple substances), as fundamental entities, with free will, is illusory and "we" are just composite objects, machines, collections of particles that interact with each other in complex yet ultimately deterministic ways (or partially random, if quantum mechanics effects are important) giving rise to macroscopic phenomena that we perceive as consciousness, free will etc. In essence, reality at the fundamental level consists of only "dead", lifeless, inanimate constituents — the fundamental particles of physics and their mechanistic interactions — while life is just a macroscopic manifestation of the aggregated effect of myriad physical interactions between these particles.

This is a very depressing concept of reality, that can be characterised as nihilistic. Nevertheless, it is promoted by its proponents as the opposite, as a progressive, enlightening, optimistic, freeing worldview. This promotion is, in my opinion, completely unjustifiable, but because of the widespread prevalence of physicalism, factors such as prejudice, intimidation, faith, complacency and others keep it from being challenged[1]. I admit that I am bemused by the fact that some people express preference for physicalism over dualism not on the basis of reason but on the basis of finding it more appealing, more desirable. For example, one can find statements such as the following:

> *"There is the lurking suspicion that the most attractive feature of mind stuff is its promise of being so mysterious that it keeps science at bay forever. This fundamentally antiscientific stance of dualism is, to my mind, its most disqualifying feature, and is the reason why in this book I adopt the apparently dogmatic rule that dualism is to be avoided* ***at all costs****. ...* ***accepting dualism is giving up****." (*[2], p. 37]*, emphasis in the original).*
>
> *"Many people insist that the complexities of human behaviour, the emotional, creative, and spiritual attributes, must be consequences of something 'greater' than physical laws. This is a wonderful concept. How much more wonderful it would be, however, if these very attributes* ***were*** *consequences of physical laws. Far from demeaning humanity, this would elevate physics!" (*[3],Chapter 18]*, emphasis in the original).*
>
> *"It is natural to hope that there will be a materialist solution to the hard problem and a reductive explanation of consciousness, just as there have been reductive explanations of many other phenomena in many other domains." (*[4]*).*

It is noted that the last of these statements is from someone who rejects physicalism on the basis of rational arguments

(of the kind examined in Section 2), yet finds it natural to hope that materialism were true instead. This shows how much the mindset of the current intellectual realm is imbued with physicalistic ideas. Although not the topic of this paper, the motivation behind the widespread adoption of physicalism deserves to be studied.

But, although it is puzzling why someone would hope that physicalism be true, it is less so why someone would believe it to be true. The success story of physics concerns not only aspects of the universe that we regard as inanimate, but also the explanation of the behaviour and functionality of the bodies of living organisms in terms of their physical structure and the physical laws, and most importantly the discovery of what seems to be physical causes of mental phenomena. Although such physical causation has been known since the dawn of human understanding (e.g. bodily damage causes pain, alcohol intoxication causes drunkenness), in recent times a much more intimate connection between brain structure/processes and mental processes has been revealed (e.g. the association of particular mental activities with neural activity in particular areas of the brain as revealed by cognitive neuroscience, or the discovery of particular brain corruptions in Alzheimer's disease). Such intimacies lend credence to physicalism because one, perhaps unwittingly, attributes metaphysical necessity to them. Why would mental phenomena need to be correlated with complex goings-on in our brains unless the latter are responsible for the former? Sure, the qualitative feel of mental experiences is patently different from the physical character and seemingly transcends the physical realm, but on the other hand this seems to be matched by the extreme complexity of the structures in our brains and of the physical processes that occur therein. This complexity must have some functionality, and what else could that be than to give rise to the mind, if the latter is so closely correlated with the physics? If the mind is a separate substance as dualists contend, then why the need for such correlations and for such physical complexity?

This is an intuitive line of reasoning that can lead one to adopt physicalism[2]. On the other hand, there are other intuitive lines of reasoning that can lead to dualism instead, which are strong enough such that the majority of people in the past and probably even today held/hold some kind of dualistic belief. One such intuition comes from the introspective examination of our own selves, and from direct observation of the mental phenomena that occur within us. Such an examination leads to the realisation that the mental realm is so distinctly different than, even opposite to, the physical one that the two cannot be one and the same. This was the intuition behind Descartes' formulation of the version of dualism that carries his name. Furthermore, for the present author the difference is not only one of quality, but also one of value: a single person seems infinitely more valuable than all of the inanimate universe combined. This intuitively leads to the idea of the immortality of the soul and the afterlife.

There is also another intuitive line of reasoning that has contributed to the endorsement of dualism, one that is not so much based on introspection, on first-person observations, but mostly on third-person observations, similarly to the intuition associated with physicalism. It seems that living, biological organisms (plants, animals, humans etc.) behave very differently from inanimate physical objects; they grow, sense, move on their own, exhibit rationality, strive for survival, reproduce, etc. The difference was particularly striking in the days prior to technological advancements that gave rise to complex machines. Due to limited scientific knowledge of how complex the physical world can be, it was not deemed possible that such complex behaviour could be due to physical factors alone. Once the organism dies, all of these attributes go away, and the behaviour of the body suddenly reduces to that of any other inanimate object. Furthermore, it decays, which suggests that there was something sustaining it, holding it together, prior to its death. In many cases, the

physical body itself does not visibly change at all at the instant of death, so it seemed plausible that the dramatic change in behaviour is due to a spiritual, immaterial ingredient being present in the body before death, and absent afterwards — the soul. On this view, therefore, the concept of the soul plays mostly an explanatory role for phenomena perceived from a third-person perspective, and misses out on the most important, first-person features of the mind. Such intuitive thinking was, I think, at least in part, the origin of views on the soul such as those of Aristotle and Thomas Aquinas, and the beliefs of the vitalists. Originating from a third-person perspective, such theories were open to confirmation or refutation by science. The evidence overwhelmingly suggests that complex behaviour such as exhibited by biological organisms can indeed arise from physical factors alone, and hence these theories have mostly fallen out of favour[3].

So, intuition can lead to conflicting conclusions, hence some or all of the above lines of thinking must be wrong. In fact, physicalism urges us to be wary of intuition, and rely only on scientific reasoning instead as the only reliable method for discovering the truth. Scientific discovery has a long history of subverting our perception of the world, to a point where we have been accustomed to this and have come to expect our intuition to be wrong. For example, temperature is the random motion of molecules, objects that appear solid consist mostly of void, the vast variation of observable macroscopic materials ultimately consist of a very small set of fundamental particles etc. It turns out that our macroscopic perception of the physical reality is illusory and what really exists is a microscopic physical world not directly perceivable by our senses. Physicalism contends that mental phenomena, their apparent transcendence, the qualitatively distinct first-person character of consciousness, are in fact just like other macroscopic phenomena: subjective, illusory and reducible to fundamental physics. In categorising mental phenomena together with the macroscopic phenomena that are reducible to fundamental physics, our everyday perception of the human person and its mental states is considered to be a folk theory ("folk psychology") that is useful but fundamentally wrong, like the intuitive folk theory that mankind held for thousands of years that the celestial universe revolves around the earth, which was also useful but wrong.

But is this applicable to the mind, and can physicalism appropriate intellect and reason for itself while deeming all other intuitions deceitful? Such questions are related to the topic of Section 3. The case of the mind is fundamentally different from all other presumably similar cases of macroscopic phenomena, because the other cases concern physical phenomena as perceived by the mind, but the mind itself cannot be a physical phenomenon as perceived by the mind. Hence the mind occupies a central and unique place. Furthermore, if intuition is something altogether illusory and deceptive, then it is not only the perception of the macroscopic phenomena that is illusory, but also that of the microscopic ones, and indeed of all of science, and physics in particular, as these are also conceived by the mind by the same intelligence, intuition, reasoning. The whole scientific enterprise has the mind and its capacity to understand reality as its source; if the mind collapses, science collapses with it. Physicalism cannot destroy its own foundations and remain standing. Our theories are often wrong but this is not because our intuition is inherently wrong (if it is, then the truth, if there is such a thing, is inaccessible to us, and physicalism, as a theory that purports to present the truth, is meaningless) but due to factors such as inadequate diligence, insufficient data, erroneous background beliefs, or even ideological bias. In my opinion, all of these factors are at play in the formation of the physicalistic thesis about the mind. Considering the data in particular, the only direct way of investigating the mind is by introspection. Prejudicial downplaying or outright rejection of this kind of investigation of the mind as illusory by physicalists leaves out indispensable data and unavoidably leads to a wrong theory. On the contrary, in order to get to the truth about the mind we must employ the full power of our

intuition and reasoning in direct introspective examination of our own selves, and in direct observation of the mental phenomena that occur within us. The full depth of the truth about the mind is therefore inaccessible to both of the third-person-perspective-based intuitive approaches described above, whether physicalistic or dualistic.
Descartes was a pioneer whose introspective intuition about the mind was exceptional, and who furthermore had the acumen to recognise the falsehood of the vitalistic intuition, despite living barely at the dawn of the modern scientific era. However, admittedly his main arguments for his version of dualism, namely the modal argument and the indivisibility argument, are rather weak.

The modal argument

The modal argument asserts that, since we can conceive of ourselves as existing without our bodies, we and our bodies are different substances. This argument is weak at best, and not even an argument at worst, depending on the meaning of conceivability, as noted by Stewart Goetz [5]. If it is taken in a weak sense as imaginability, as it is meant by modern proponents of the argument, e.g. [6][7], then the argument is of very limited value, and in fact seems like a recipe for self-deception. We know very well that we can imagine things that seem possible at first glance but turn out to be impossible if we look closer. For example, I can imagine travelling faster than the speed of light, but the theory of relativity says otherwise. Hence, the physicalist may reply that although I can naively imagine my existence apart from my body, if I studied precisely the inner workings of the brain and saw that consciousness is deducible from the brain's structure and functionality according to physical principles then I would realise that it is impossible to have consciousness without a brain, or at least some other physical system of similar functionality. To support dualism, a stronger kind of conceivability is needed, one that is synonymous with logical (metaphysical) possibility, i.e. there must be no logical barriers for my existing without a physical body, which requires that there is no way to reduce consciousness to something physical. But then the modal "argument" is not really an argument but merely a statement of the truth of substance dualism; the real issue would be why my existence without a physical body is strongly conceivable, about which the "argument" is silent.
What I think is going on here is that Descartes, and possibly also the modern proponents of the argument, intuitively, through introspective examination that they describe as "imagining themselves without their bodies" but involves more than that, realise that the self/ego/person — what it is that they fundamentally are — is so sharply different than anything physical that it cannot be physical. This is ultimately an exercise that everyone should take on themselves to gain an understanding of what they really are. Fundamental truths are ultimately accessible only through intuition directly, not by way of argumentation; even our most rigorous discipline of mathematics is necessarily based on self-evident, unprovable axioms. The truth about the nature of the self is, in my opinion, fundamental, and ultimately accessible only through intuition and introspection. The role of arguments is to reduce the intuitive distance that one has to travel to get to the truth, but they can never eliminate it completely; one cannot avoid doing some hard work to get to the truth. In the case of the self the truth is quite deep, which makes it easy for people to willingly refuse to travel the distance needed, and hence I do not expect that arguments for substance dualism can swiftly change the balance in its favour, although eventually the truth will be accepted. When it comes to the modal argument, the "weak" version is of little help, while the "strong" version merely points one to the right direction and asks him/her to travel all the distance by themselves.

More useful, especially for a modern person accustomed to scientific thinking, are arguments that instead of claiming that it is possible (conceivable) that the mind is non-physical, claim that it is impossible (it is inconceivable) that it is physical. A group of such arguments are at the core of what David Chalmers has termed the "hard problem of consciousness" [8], which is the topic of Section 2. It is formulated in terms of consciousness as a non-physical property and leaves open the issue of whether the substance that exhibits consciousness is an immaterial substance (a mind, person, self etc.) or something material (property dualism). This group of arguments includes a modal argument that is the reciprocal of the Cartesian modal argument; while the latter claims that it is conceivable that there exists a mind without a body, the former claims that it is conceivable that there exists a functional body without a mind (a concept known as "philosophical zombie"). The kind of conceivability commonly implied in the argument is a relatively weak one, and hence while the argument has been influential, it has not been decisive. However, in Section 2 it will be argued that a strong conceivability version of the argument is much easier to support than for the Cartesian modal argument, and that in fact strong conceivability is the natural sort of conceivability for this argument.

The indivisibility argument

The other Cartesian argument is the indivisibility argument, which contends that since the mind is simple (indivisible), while the body is complex (divisible), they cannot be one and the same. For example, in the 6th Meditation of the "Meditations on First Philosophy" one reads:

> *"[T]here is a great difference between the mind and the body, inasmuch as the body is by its very nature always divisible, while the mind is utterly indivisible. For when I consider the mind, or myself in so far as I am merely a thinking thing, I am unable to distinguish any parts within myself; I understand myself to be something quite single and complete. Although the whole mind seems to be united to the whole body, I recognise that if a foot or arm or any other part of the body is cut off, nothing has thereby been taken away from the mind. As for the faculties of willing, of understanding, of sensory perception and so on, these cannot be termed parts of the mind, since it is one and the same mind that wills, and understands, and has sensory perceptions... This one argument would be enough to show me that the mind is completely different from the body, even if I did not already know as much from my other considerations."* [[9], p. 119—120]

This argument also has modern proponents, e.g.[5]. But the physicalist need not contend that mind and body are one and the same; rather, he/she can assert that consciousness is a physical phenomenon caused by the body (brain). Similarly, the luminosity caused by a lamp is not identical with the lamp, nor is the music generated by a digital audio player the same as the device that generates it. Just like cutting off one of my arms does not reduce my mind, the luminosity and the music will not be reduced if one cuts a piece off the lamp's base, or a piece of plastic off the player's casing. But if one cuts a vital component, e.g. a wire, then the luminosity and the music will cease — but the same will happen with my consciousness if someone cuts off my brain or even my heart or some vital organ. In this physicalist line of thought the body is often described as hardware that implements the software that is the mind.

So, I do not think that Descartes' argument, as it is phrased above, takes us far. It may, though, help to dispel the "folk"

theory that many hold, perhaps unwittingly, that we are identical with our bodies or that we are at least partly constituted by our bodies (we are a composite of mind and body). This "theory" emerges naturally due to the intricate interdependency of mind and body and is very helpful, convenient and practical for everyday life. But the notion of a human body, which we intuitively assume to be an objective entity (and of a person as a composite of body and soul), is in fact subjective, a human convention, just like with any other physical macroscopic object like a chair, a phone, a mountain etc. — see comment [7]. On the contrary, I think that Descartes points our attention to a more fundamental sense of self, one that exists in an objective/absolute and not subjective/relative sense. We are an indivisible centre of consciousness, of first-person perspective, of existence I would say, and this is something real, not by convention. The "indivisibility" or "simplicity" that Descartes is referring to here is that pertaining to what in philosophical terminology is called a substance. Nevertheless, he was not able to express his intuition fully and the resulting argument fails to convey it. However, there are arguments that can help one delve deeper into the nature of the self/person/ego (whatever one wants to call it) and understand that it is a simple substance, and these arguments are the subject of Section 4, which in my view is the most important part of this essay.

It should be noted that the aforementioned physicalistic understanding of consciousness as a physical phenomenon is fallacious, because phenomena are perceptions of the mind. The luminosity we see and the music we hear are mental phenomena that occur in our minds only; similarly, the output of a computer software has meaning only for a mind. From a physical point of view, the lamp emits photons and the audio player creates pressure oscillations in the air, which are physical events of a completely different character than mental events. These physical events cause a chain of other physical events that arrives in our brains, and from there somehow the mental experience ensues. That same mental experience could in principle be generated without a lamp or without an audio player (e.g. by electrodes in our brains — or possibly directly in our minds?). So, luminosity and music are mental phenomena experienced by a mind, and hence cannot serve as an example of what a mind is; and their physical counterparts of photon and pressure wave emission are obviously not candidates of examples of what a mind is, being completely dissimilar to it. These issues will be explored further in the sections that follow.

So, let us begin our exploration of whether it is dualists that are like people who believe in ghosts, magic, fairies, dragons and the like, or if it is physicalists, trying to explain consciousness through physics, that are like medieval alchemists trying to chemically transmute lead into gold, or like ancient geometers trying to square the circle. In the process, we will examine also other, intermediate theories such as panpsychism.

2. The physical inexplicability of consciousness

2.1. Introduction

This is the original "hard problem of consciousness", so named by Chalmers[8]. It is a hard problem for physicalism, and has been highlighted through several arguments such as the modal argument about the conceivability of "philosophical

zombies" (things that are physically identical to humans but lack consciousness) [10], the "knowledge argument" (knowledge of all physical truths does not imply knowledge of all the truths about consciousness) famously demonstrated with the imaginary tale about Mary the neuroscientist [11] (see also [12]), and the "explanatory gap" argument according to which consciousness is not explicable by physical principles [13] — in fact, Leibniz' "Mill" argument belongs in this category as well. According to Chalmers [4], all of these arguments have in common that they demonstrate the existence of an epistemic gap between physical and phenomenal (mental experience) truths, which suggests that there is an ontological gap as well. However, I would argue that the ontological gap does not merely follow from the epistemic gap; rather, conversely, in order for one to admit that there is an epistemic gap one has to first admit, perhaps tacitly, that there is an ontological gap. All of these arguments demonstrate that mental experiences, qualia, are not explicable by (or deducible from, or implied by) physical principles. In order to follow the arguments, one must introspectively examine the nature of his/her mental life and experiences and honestly admit that they are so qualitatively different from the physical realm that deducing them from physical principles is impossible (or, in the case of the modal argument, which is the weakest of the three[4], that their independence from physical principles is conceivable). This may sound like assuming the conclusion ("question-begging"), but in reality it is just a sincere observation. One cannot examine something without observing it and taking notice of its characteristics; in the case of mental phenomena, their most distinguishing characteristic is this first-person aspect of them, whose qualitative character is so patently different from anything physical. One cannot arrive at any reliable conclusion about consciousness if he/she chooses to disregard this characteristic. In fact, it is often physicalists who assume the conclusion, downplaying or outright denying this distinct qualitative character of mental experiences precisely because it does not fit into the physicalistic framework.

In brief, the crux of this hard problem is that the principles of physics (and of higher-level sciences derivable from physics) do not entail the existence of mental phenomena, but in fact they are qualitatively disjoint from them. Thus, reality is not entirely physical. I will try to explain this by the following line of thought, which is similar to the "explanatory gap" argument, but places emphasis on the ontological rather than on the epistemic side of things.

2.2. Attempting to deduce consciousness from physics

Physicalism claims that all of reality is purely physical at the most fundamental level. A very important aspect of reality is mental phenomena, the subjective experiences of conscious beings, also called "qualia" in the language of philosophy, and often collectively referred to under the term "consciousness". If physicalism is true, then consciousness should be deducible from physics. Of course, physicalism has to concede that some elements of reality are primitive, inexplicable, fundamental, such as perhaps time, space, fundamental particles, and their properties and the laws that govern them; the explanatory cascade from complex entities down to simpler ones cannot be infinite, but a bottom must be reached at some point. But a fundamental tenet of physicalism is that this set of primitive components must contain only inanimate, "dead" elements that behave mechanistically. Thus, it is crucial for physicalism that consciousness be explainable in terms of physical quantities and laws, rather than being itself primitive.

However, how could a system made of physical elements that are required to lack consciousness, and which interact according to laws that are also required to lack any reference to consciousness, produce consciousness? A naive answer

might be that a lot of things appear from a macroscopic perspective very different than do their microscopic constituents, so why not consciousness? But the key word here is *"appear"*: they appear so to sentient beings, to bearers of consciousness; while that may be true for macroscopic phenomena, it cannot be true of consciousness itself, because in order for this to work, consciousness is a prerequisite, so it cannot be a result. We cannot explain consciousness in terms of consciousness, circularly. If then we set aside how macroscopic phenomena look and feel, which is subjective (and is in fact the subject of our investigation, whether the way we perceive things is physically deducible or not), as far as we know the objective behaviour of macroscopic physical systems is determined entirely by the physics of their elemental microscopic constituents[5]. By considering the multitude of microscopic elements, their precise initial locations, and the microscopic laws that govern their behaviour, we can, in principle, predict precisely how the macroscopic system will evolve, either by itself or under external influences.

Thus, science has a layered structure, where each layer corresponds to the physical reality as viewed on a different spatiotemporal scale. Higher-level entities and properties are defined in terms of aggregates or statistical averages of lower-level entities (e.g. bundles of atoms held together by covalent bond forces are called molecules, and the density of a material at a point is the statistical average over time of the total mass of the molecules contained in a small volume around that point, divided by the volume). The physical laws of the higher-level sciences are deducible from the low-level structure of its entities, quantities and properties, the low-level physical laws, and a statistical averaging procedure. Thus, for example, chemistry is deducible from physics, and biology from chemistry. Historically, however, sciences of all levels have been developed more or less in parallel rather than sequentially, so that higher-level laws were usually discovered empirically rather than by deduction from lower-level sciences, and the deduction was performed, if so, at a later stage. Often these empirical macroscopic laws are only approximate, especially when an accurate statistical averaging of the microprocesses does not result in a neat, simple, tractable macroscopic mathematical law. However, in modern times it is becoming increasingly possible, due to advances in computational technology and algorithms, to predict phenomena at a relatively coarse scale (higher-level phenomena) directly from laws and processes that occur at a lower level, through numerical simulations that can account for a very large number of micro-constituents and their interactions.

Consider for example the water molecule, H_2O. It consists of two hydrogen atoms bonded to an oxygen atom. The structure and properties of the water molecule can be deduced from the structure and properties of the hydrogen and oxygen atoms, which we know from quantum mechanics. Then, the behaviour and macroscopic properties of a large collection of H_2O molecules, such as a water droplet, a snowflake, or an ocean, can be similarly deduced. Macroscopic properties such as density, phase (gas, liquid or solid), temperature, colour, transparency, viscosity, surface tension etc. can be deduced from microscopic properties such as structure, mass and charge distribution of the H_2O molecules, their velocities etc. The macroscopic behaviour of water can also be deduced, e.g. how a mass of water will flow or deform under the action of an external force; this can be performed either by solving the Navier-Stokes equations, which are empirical macroscopic equations derivable from microscopic molecular considerations, or by directly simulating the motions of the huge number of H_2O molecules comprising the water mass via molecular dynamics simulations. More useful simulations are routinely performed about the structure and functionality of proteins, e.g. simulation of the behaviour of the spike protein of the SARS-COV-2 virus when in proximity to the ACE2 receptor of our cells may reveal how the virus enters our cells and which virus variants are more likely to be more successful and dominant[14].

So, what is the point of all this discussion? The point is this: if consciousness is to fit into the physicalistic framework, it must be deducible from physical principles, since if it is fundamental then physicalism is false. We don't have to deduce it directly from fundamental physics, any higher-order entirely physical science or combination thereof will do (physics, chemistry, biology etc.) — it will then be automatically deducible from fundamental physics as well. But merely discovering empirical laws (correlations) between mental phenomena and brain structures and processes does not suffice, since this would not rule out that consciousness is fundamental; such correlations are also compatible with panpsychism and other, stronger versions of dualism, even with versions of idealism. So the question is: is consciousness deducible from physics? A naive physicalist reply may be that it is, but we haven't yet deduced it because the functionality of the brain has not been deciphered yet due to its great structural complexity. In other words, consciousness is a high-level phenomenon, but the low-level structures and processes of which it consists are combined in such a complex manner that we have not yet been able to theoretically deduce the high-level (consciousness) laws and phenomena from the low-level (physics, chemistry, biology) ones. But if we are unable to perform the theoretical deduction from micro- to macro-level, we are still left with the brute-force option to perform predictions of the macroscopic phenomena (consciousness) directly from the microscopic phenomena by numerical simulations.

So, let us suppose that instead of modelling a single protein, we perform an atomistic molecular dynamics simulation of a whole human body. Of course, with the present state of the art in computer technology such a simulation may require billions of years to complete, but let's assume that we either happen to have computers much more powerful than modern ones in our disposal, or that we have the time to spare. What would we hope to find from such a simulation? Well, the same kind of information as when modeling a single protein, albeit at a much grander scale. The kinds of results we can hope to obtain are dictated by the equations we solve; these equations involve time, locations, velocities and accelerations of particles (or probabilities thereof in the case of quantum mechanics), their masses and charges, etc. — all of them physical quantities, nothing directly related to consciousness. So, similarly to a protein, the simulation will give us a prediction of the evolution in time of the configuration of the human body, tracking every single molecule, atom, even electron. It will be able to predict macroscopic tasks ranging from mechanistic ones such as breathing and pumping of the blood by the heart, to ones that are considered intellectual, like thinking. But, concerning the latter, what exactly would be simulated? Obviously, only the physical aspect, i.e. the chemical processes in the brain, the motion of ions, atoms and molecules. For example, suppose we simulate someone reading a trigonometry book and learning about the Pythagorean theorem. Our simulation would include light rays reflected on the book page and arriving at the body's eyes, where neural signals would be generated and travel to the brain; there, some structural changes in the neural networks of the brain will occur, as well as biochemical changes within individual neurons, reflecting the processing and storage of the newly acquired knowledge. Then, when the page with the homework exercises is reached, the optical signals generated due to this page facing the eyes will initiate processes in the brain which will ultimately result in signals traveling down to the muscles of the arms and hands to make them pick up a pen and draw lines on a paper which, to a sentient being, will represent the solution to the exercise. Obviously, just like when simulating a protein or a bunch of water molecules, none of this entails the manifestation of consciousness. All the simulation can give us is the temporal evolution of the locations and velocities of each atom of the body, and of the associated force fields and energy potentials.

If physics provides a full description of reality without leaving anything out, then our simulation tells us all there is to know; and the simulation results do not include consciousness, hence there should be no consciousness whatsoever. Our simulated body, if it existed in reality, should function just like a "philosophical zombie", or a Descartes' "automaton" (a better term, in my opinion, might be "biological robot", a robot made of biological rather than mechanical and electronic technology). Thus, the fact that we happen not to be zombies or biological robots, as each of us knows from personal experience, disproves physicalism. Of course, it may be argued that we have not discovered all of physics yet; but even so, for the spirit of physicalism to be true, any newly discovered foundational element of reality would need to be equally inanimate, dead, as the ones we currently know of, i.e. force, mass, charge etc. Therefore, the associated new laws, if included in our simulations, would still fail to predict any consciousness at all. Physics does not, and can not, provide a complete description of reality.

2.3. Supervenience

A possible objection is that physicalism may be true without consciousness being deducible from physics if physicalism is defined in terms of *supervenience* [15]. According to this more general definition, physicalism would be true if consciousness is supervenient on the physics of the body, meaning that it is necessitated by certain physical bodily arrangements and processes, even if it is not deducible from them. According to Davidson [16],

> *"... mental characteristics are in some sense dependent, or supervenient, on physical characteristics. Such supervenience might be taken to mean that there cannot be two events alike in all physical respects but differing in some mental respect, or that an object cannot alter in some mental respect without altering in some physical respect".*

First of all, let it be mentioned that it will be argued in Sec. 4.1. that it is indeed possible for there to be very significant mental differences while the corresponding physical structures are exactly the same (two different persons having exactly the same bodies, molecule for molecule). Therefore, physicalism is false even if thus defined. Furthermore, the effort to remove the deducibility condition but maintain the necessity condition is suspiciously strange, giving the impression of an attempt to sweep the problem under the rug. After all, the momentum of physicalism has been built on the explanatory and predictive power of physical sciences, and here we have an admission of the failure of this power in the case of mental phenomena. The above definition claims that the mere existence of necessary correlations between the mental and physical is sufficient for reality to be characterisable as physicalistic. But obviously, the nature of the necessity of these correlations is crucial for whether the spirit of physicalism is maintained or not. In my opinion, only if this necessity means "follows logically from physical principles" is the spirit of physicalism maintained. In any other case, even if the supervenience version of "physicalism" were nominally true, in essence physicalism would be false.
For example, consider the case that the correlations between mental and physical phenomena are nomological, there is no reason behind them. These correlations could then be formulated into a set of so-called psycho-physical laws that are primitive, fundamental, inexplicable[6]. Although physicalism, defined in terms of supervenience, would be nominally true,

the spirit of either substance dualism or panpsychism, depending on whether the mental phenomena that occur in our minds are themselves primitive (they are the ones that appear in the primitive psycho-physical laws) or are analysable into micro-mental phenomena (the psycho-physical laws are of the form posited by panpsychists) would be more representative of reality. Note that the present case is not about there being a hidden connection between mental and physical phenomena, one that exists but is epistemologically inaccessible to us, but about there really being no connection at all (as awkward and unlikely as this sounds especially in the case that our mental phenomena are primitive and not analysable into micro-mental phenomena) other than that they always appear in tandem. In the absence of any palpable connection between mental and physical phenomena there are no real grounds to support the spirit of physicalism that reality is entirely physical. Any argument to support this would be sophistic.

Or, consider the scenario where minds and bodies are separate substances but God has interwoven them in this life (as opposed to the afterlife) because this is beneficial for the maturation of minds. That is, the aforementioned mental-physical correlations are sustained by God because he has a purpose; the way the correlations work is miraculous, is not explainable in any way, since God and his power are transcendent. For all beings other than God the correlations are in a sense necessary, since God is omnipotent and nothing can go against his will. This scenario should clearly be ranked among dualistic worldviews, but oddly enough would also qualify as supervenience physicalism.

It follows from the above discussion that supervenience physicalism is too weak and vague to count as real physicalism, and therefore it has lost ground [15].

Finally, the possibility has been proposed (e.g.[11]) that there is indeed a physical connection between mental and physical phenomena, but it is epistemically inaccessible to us. That is, it is not that we have not yet discovered it, but we can never discover it due to epistemic limitations that we as humans have, for example because there has not been any evolutionary pressure for us to acquire this epistemic ability. This argument, however, shakes the foundations of physicalism rather than supports it. If the deducibility of the mental from the physical seems impossible and yet is true, then also the deducibility of other physical phenomena from fundamental physics may seem true but be false, because evolutionary processes have resulted in things falsely appearing to us this way. In this case, our logic cannot be trusted and we can never know whether physicalism, dualism, or something else is true, no matter what our reason tells us. Physicalism is a position that claims to describe truth, but if truth is inaccessible to us, then what use is it to posit physicalism?

2.4. Property dualism and panpsychism

The non-deducibility of mental phenomena from physics leaves open the possibility for a version of materialism where physics does not tell us everything there is about reality. Maybe there are other, non-physical, mental qualities in matter, and in its fundamental elements in particular. If we could formulate laws that include properties of this kind, then maybe our simulations would be able to predict consciousness. Such a theory is appealing to persons with a physicalistic mindset who nevertheless have the acumen and candour to recognise at least some of the hard problems of physicalism (usually the present one, which is the easiest for a physicalist to grasp). It tries to salvage what it can from physicalism, keeping the mindset as intact as possible, while overcoming the explanatory gap. In particular, it ignores the carrier of consciousness, the person, which it considers, like physicalism, to be just a complex phenomenon arising from the

combination of processes (physical and mental) at the microscopic level, and not a real entity, a substance. Thus, the foundations of the physicalist worldview, at the substance level (i.e. fundamental particles), remain relatively intact, while consciousness can be downgraded to another kind of property of the fundamental material constituents of reality, like mass or charge, only of different qualitative character.

In order to study anything and get to the truth about it, there is no other way but to observe it carefully and use intellect and intuition. This is how science has borne fruit when it comes to our understanding of the physical world. When it comes to mental phenomena, there is no other direct way to observe them but introspectively, which everyone can do without the need for expensive laboratories and equipment. However, the prevalence of physicalism has resulted in the introspective observation of the mind being regarded as deceptive, illusory, and access to the truth about it being attempted mainly by third-person observations of the brain and of behaviour. Property dualism stems from a partial understanding of the discrepancy between mental and physical, achieved through some introspection and intuition. But the proposed solution to the hard problem seems lazy and retrogressive to a mindset that again avoids the introspective dive into the nature of the mind, the self. Hence, property dualism is not immune to the sort of hard problems that will be presented in Sec. 4 and pertain to the carriers of consciousness, persons.

A notable theory that may or may not be classified as property dualism is panpsychism[17][18]. Panpsychism developed from an intuition of some philosophers that the physical laws do not tell us everything there is about matter, because they only describe its behaviour from a third-person point of view, extrinsically, whereas it must also have some intrinsic nature than cannot be discovered by third-person observation. In their effort to fathom what such an intrinsic nature could possibly be, they could come up only with consciousness — the first-person perspective itself. Thus, panpsychism purports to reconcile the seemingly contrasting mental and physical natures by regarding them as two aspects of the same single substance that is matter.

But, while on the surface it may seem this way, I think that the case for panpsychism is almost as problematic as that for physicalism. What makes physicalism implausible is the patently dissimilar characters of the mental and physical aspects of reality, which precludes the possibility that mentality could have a physical origin. Although panpsychism acknowledges this conspicuous qualitative difference, its claim that mentality and physicality are two aspects of the same thing still seems implausible given the width of the gap. Two aspects of the same thing could be different in various ways, but there must be similarities too, different reflections of the features of the same underlying essence. The mental and the physical seem to share nothing in common, their natures seem completely unalike, disjoint, irreconcilable, even diametrically opposed in certain respects. Mentality has to do with reasons, rationality, understanding etc. while physicality only with mechanistic behaviour governed by mathematical laws involving space, time, mass, forces etc. (these issues will be discussed in Sec. 3). Behaviour based on reasoning and behaviour based on mechanistic principles cannot be one and the same thing. According to one panpsychist view [19], panpsychism inverts the assumption often implicit in physicalism that consciousness is like software and the brain is the hardware which implements it; for panpsychism, consciousness (what matter really, intrinsically, is) is the hardware, and the physics of matter is the software that is implemented by the hardware. In this case, panpsychism does not avoid the hard problem discussed in this section, but merely inverts it;

whereas for physicalism the problem is that the derivation of consciousness from physical principles alone is strongly inconceivable, for panpsychism what is inconceivable is the derivation of physics from consciousness. In both cases this is due to the qualitative chasm between the mental and the physical.

Although panpsychism assigns greater value to the mental, regarding it to be the intrinsic nature of matter — what matter really is — it actually deprives it of any role in determining either physical or mental phenomena. Since panpsychism endorses physical causal closure, it is a form of epiphenomenalism where mental phenomena cannot have any effect on our behaviour. If panpsychism accepts that there is a complete mapping between someone's mental and physical states, i.e. a form of weak supervenience, and since the physical state (the brain and its internal workings) is governed solely by the laws of physics, then ultimately mental phenomena are also governed by the laws of physics, and not only can they not have any effect on the body, they also cannot have any effect on the mind itself. A panpsychist may object that, according to his/her theory, the laws of physics have a mental origin and should not be seen as contrasting with mentality. But this argument is not convincing, since mentality has to do with reasons, meanings, understanding, judgement, choice etc. whereas the laws of physics, which panpsychism acknowledges as ultimate governors of the functionality of our bodies and hence of our behaviour, have an inanimate, mathematical, mechanistic, fixed, unconscious character. If panpsychism is true then mentality and its aforementioned traits are reducible to simple mathematical formulae.

Another murky point for panpsychism is that, whereas it associates mentality with the essence of matter, scientific experience associates it with the structure and functionality of matter instead, and not with matter itself directly: the reason why consciousness is exhibited by a brain and not by a rock, say, must have something to do with the structure and functionality of the brain. If the same matter that is organised into the brain is converted into a pulp, consciousness will be lost. In fact, even much smaller changes in the structure of the brain can bring the loss of consciousness. But panpsychism claims that consciousness is the inherent nature of matter, not of the structure of matter. Why then is there the need for the extremely complicated neural machinery of the brain?

On the face of it, physicalism and panpsychism appear to offer opposing views of reality in the sense that, whereas according to physicalism everything is fundamentally dead, according to panpsychism the matter of which everything consists has, fundamentally, some sort of life. But, in my opinion, the difference between them, while significant in principle, is insignificant in practice; since going down the biological complexity ladder from humans, to mice, to insects, to bacteria etc. it is reasonable to assume a corresponding reduction in the conscious capacities of the organisms, it follows that when we reach the fundamental level of elementary particles of matter the (proto)consciousness of these will be negligible. So, in panpsychism the elemental particles of matter are essentially exactly the same as in physicalism, except that they additionally have the mere *potential* to produce consciousness when combined appropriately into complex structures. To me, panpsychism has almost the same loathsome odour of lifelessness as physicalism, the belief that life, consciousness, existence, are macroscopic phenomena that are built from "dead" foundational components, whereas intuitively it seems that it is they that should be the foundational components of reality.

For panpsychism, the human person, such as me and you, remains a composite object, an illusion, like in physicalism, the difference being that the ingredients include also the (proto)consciousnesses of the particles that make up our bodies.

While the (proto)consciousness of an elementary particle is recognised as something real, fundamental, primitive, it is not so for *our*, macroscopic, consciousness, which is regarded as just an aggregate that our subjective perception deceives us into regarding as a unity. On this point in my opinion panpsychism exhibits a lack of intuition and introspective contemplation and follows an "easy" path of abstract thinking along the lines of physicalism. In Sec. 4 it will be argued that the person is fundamentally simple, and hence panpsychism is vulnerable to hard problems involving persons. A notable related problem is the so-called "combination problem" of how a unified human consciousness can result from the combination of the simple consciousnesses of the particles of the body [20].

Nevertheless, I do believe that there is some truth in the panpsychists' intuition. They attributed consciousness to matter because they thought that for something to truly exist it must have an intrinsic nature, and acknowledged that the only possible kind of such nature is a conscious one. In my opinion, in order for something to truly exist, in a complete and perfect sense, it must be conscious, it must be a center of first-person perspective. Existence, life, and consciousness are in a sense synonymous. But that does not mean that other things, that lack consciousness, cannot exist in a relative sense. Such things are nothing in and of themselves, but acquire value and existence from the effect they have on minds, on real existences. I regard matter as belonging in this category.

3. Meaning and the physical language of thought

The hard problem of the previous section is usually discussed in relation to sensory conscious experiences, e.g. seeing red or feeling pain. This is somewhat convenient for the physicalist, because it portrays a picture where the only problem for a physicalistic interpretation of the mind is this special qualitative character of these experiences, why there is something that having them is like. It is implied that all other, more important, aspects of the mind are, in principle, obviously explainable in terms of physics, in terms of the structure and functionality of the brain; that the sensory qualia are mere mysterious byproducts of the functionality of the brain which do not play any central role in the functionality of the mind.

But is this the case? Our mental faculties go much deeper than sensory experiences. Usually, the latter are starting points for deeper processes where the mind represents and understands the reality in which it exists. In fact, it seems to me that even if a mind completely lacked any sensory faculties, i.e. it was isolated from the outside world, it could, by introspectively observing its inner self, form some sort of idea about the world and reality. After all, as Descartes noted, our own existence is the only thing that we can be absolutely certain of. Nevertheless, the fact that I am not completely autonomous, I am neither the source of my own existence nor the sole master of my fate, and, while I can exert some limited influence on my environment, the environment has a greater influence on me, means that there is a reality that extends beyond me, a world of which I am but a part. The ability to think and try to understand both one's own self and the external reality, to consciously perceive elements of reality as such, are defining characteristics of a mind, of a "thing that thinks" as Descartes said. Minds are rational *observers* of reality (and also *actors*, agents affecting/shaping it by exercising their free will, but this will be the topic of Sec. 5), they have conscious and rational access to it (direct, in the case of their own selves, indirect in the case of the external reality), however incomplete, limited, distorted and faulty. The meaning of

"thinking" here is not restricted to logical reasoning, but encompasses all mental activities that have to do with reality, including desiring, hoping, fearing, intending, judging, loving, hating, etc.; all of these have to do with aspects of reality. Even imagination, thinking/inventing of fictitious things and states of affairs, has its source in our perception of reality, bears a deliberate and unavoidable semblance to it, presenting an alternative version of it, collaged and inspired from perceived elements of reality itself (but also perhaps from unperceived innate concepts).

It is reasonable to be an anti-realist concerning some aspects of our perception of reality[7], but full-fledged skepticism is incompatible with physicalism, since the latter holds that reality exists and is physical, and that we can know this based on evidence and reason. Therefore, physicalism requires at least a minimum threshold of commitment to realism, and to the belief that we have epistemic access to reality. Furthermore, it has to be accepted that we have the ability to understand reality, an innate faculty we call rationality that allows us to see how reality "works", know what is possible or probable and what is not, make predictions, think of possible explanations, evaluate them, etc. These are the foundations upon which the edifice of science has been built. Like the qualitative sensory experiences, thinking, understanding, making sense of reality etc. have a distinctive, first-person, mental character which contrasts with the unconscious, inanimate, mechanistic character of the physical realm. The former character can be said to be the mark of the mental, whereas the latter character can be said to be the mark of the physical. The project of physicalism is to show that mental phenomena can be reduced to physical phenomena, i.e. that reality at its foundations has the physicalistic, unconscious, mechanistic character, while the mental character is something that appears this way from a macroscopic and subjective point of view, whereas if we look at it more closely and attentively it will be understood to be nothing more than complex physical, inanimate, unconscious processes.

The problem then is that it simply is not intelligible how the genuine understanding of reality that we mentally experience and engage into, which is a requisite for conceiving physicalism and science, is just mechanistic motions of molecules and ions inside our brains, whatever pattern these motions may follow; that it merely amounts to electromagnetic forces pulling and pushing atoms, molecules and ions according to physical laws of simple mathematical character that have nothing to do with the meanings that are the objects of our thoughts and our understanding. Physicalism would imply that our very concepts of space, time, molecules, mass, forces, charges, mathematics, physical laws etc. are explicable by the very objects that these concepts refer to. In other words, metaphysics, which seeks to find the meaning and essence of the physical world (or a wider world, if physicalism is false) would itself be explainable via physics in a reversal of roles; metaphysics would therefore ultimately be redundant and physics would be able to explain and interpret its own self. But the likelihood of this already sounds quite weak, since it seems implausible that a closed set of laws or rules can be such that they follow from the rules and laws themselves, that they are self-derivable, as would be the case if meanings and concepts are something physical.

The present hard problem is therefore an extension of the previous one, with more serious and deep repercussions. Physicalists who try to explain the mind overlook this problem by anawarely embracing a sort of dualism: they regard themselves as observers in the full sense given above, with full access to reality and unrestricted capability to understand the truths about the mind, while their subjects of investigation, which are as human as they themselves are, they regard as physical systems whose apparent mentality is something illusory, commonly perceived in a "folk-theoretical" manner, subject to limitations by, and conformance to, the underlying physics to which it is reducible, and they (the physicalist

philosophers, using the same illusory mental capacities that their subjects have) can help us see clearly the actual truth behind it. Various aspects of the hard problem of meaning and mental content are discussed in what follows.

3.1. Intentionality

The present problem is often referred to as the problem of *intentionality*, a term that indicates that mental states are *about* things, properties, or states of affairs; our thoughts have objects towards which they are directed. The term was coined by Brentano [21] and made popular by Chisholm [22] both of whom famously held that intentionality is the mark of the mental. The present work avoids this term because in my opinion it is restrictive and does not do full justice to the mind. Our thoughts are not merely directed at objects, but most importantly involve conscious, first-person experiential understanding of actual or possible elements of reality; they are associated with our consciously accessing reality itself, "touching" it, consciously understanding it, something whose character is completely alien to the physical. Speaking only of "intentionality" or "aboutness" allows physicalists to draw parallels with purported physical manifestations of these features, such as the directedness of a compass towards the north pole, the indication of smoke that there is fire, or a thermostat's aim towards a certain temperature [23][24], and claim that it is the same sort of intentionality that we are talking about in both cases (mental and physical). This is obviously to confuse a physical phenomenon (e.g. the alignment of a compass needle with the north pole by magnetic forces) with an understanding of that phenomenon. In a physical interaction, one physical entity is influenced by another; by understanding the mechanics of the interaction we can make inferences for one entity by observing the other. In the physicalists' view, conscious understanding of the interaction and the interaction itself are one and the same thing.

Some acknowledge a difference between the intentionality exhibited by the objects of the simple examples such as the compass or thermostat and that exhibited by humans ("natural" versus "non-natural" intentionality/meaning [25]), but regard it as quantitative, a matter of degree of complexity [24]. According to this view, a complex enough physical system would exhibit the same intentionality as a living person, and the required complexity could certainly be achieved by an evolutionary process driven by natural selection. Thus, complexity and natural selection are used in place of "God" in "God-of-the-gaps" reasoning, choosing to overlook the fact that the difference is ontological and not merely a matter of complexity. Other highly implausible and desperate attempts to propose a continuous transition from natural to non-natural intentionality (making "non-natural" intentionality essentially natural) will be discussed later.

It is surprising that it is relatively recently that Horgan and Tienson[26] made a significant impact by noting something fairly obvious, namely that human intentionality, the kind that is mysterious and of interest, is always inseparable from "phenomenology": thought is both intentional and phenomenological i.e. a conscious, first-person experience. But even this characterisation of thought does not do it full justice: a phenomenal (conscious) experience can be something as simple as seeing red or hearing a sound; and intentionality simply means being "about" something. Thought goes much deeper than that as it involves *understanding,* our conscious accessing of reality, which is an unfathomable mystery. Analyzing thought into the two components of intentionality and phenomenology is a crude simplification that can give the false impression that it can be analysed just like any external physical object when in reality thought is the very thing that we use to analyse and understand everything including itself.

3.2. The indeterminacy of representations of reality in physical terms

Just as conscious access to reality, understanding of it, is a mark of the mental, lack of this capacity is a mark of the physical. How, then, could physicalism be true? How could minds ultimately be physical at the fundamental level? This would require that conscious reasoning and understanding can be implemented by a physical system. But while we can construct systems whose function superficially resembles the structure of thinking, e.g. computers, the real understanding is done by people, minds, who assign meaning to the meaningless physical states inherent in, and physical operations performed by, these systems.

For example, the functionality of computers is based on their ability to handle two basic physical states that we denote as "0" and "1" and to perform very basic physical operations on them; all information we may wish to represent in a computer is coded by sequences of these two states, be they integer numbers, real numbers, colours, pixel coordinates, musical symbols or sounds, letters of the Latin, Greek, or Chinese alphabet, stock prices, novels, pictures, movies or whatever other data we can think of. The computer has the built-in capacity to process sequences of "0"s and "1"s in specific ways, and the programmer provides these sequences and the order in which the operations on them are to take place (the computer program). The computer has no understanding of what the "0" and "1" sequences represent or mean, and mechanistically operates on them based on the instructions, in completely deterministic physical processes governed by the laws of electromagnetism[8]. Now, these sequences by themselves have no specific meaning, but we can assign meaning to them by mapping them to elements of reality according to some convention. For example, the same sequence of bits 00100110 could represent the integer number 38, the character "&" in the ASCII system of characters, a particular sound, a specific shade of grey, or countless other things. What it represents is not inherent in the sequence 00100110 itself, but is decided upon by a mind. Thus, the mapping of physical elements in the computer to meanings of elements of reality is *indeterminate*. It does not follow deterministically from objective logical rules, but requires a subjective, mind-invented convention; a vast number of such conventions are possible, each giving rise to a different mapping, so that each bit sequence can be assigned an vast number of meanings.

Similarly, the most basic physical means invented for the communication of thought is language, which represents meanings via sounds or geometric marks on paper or a screen. All human languages, including languages in an extended sense such as the languages of mathematics, of music, of computer programming etc. are by convention: a signifier (the physical element) is chosen to correspond to a signified (what is meant). This correspondence has been invented by minds, and agreed upon between minds, whereas it does not have any determinate objective basis; there is no "hard" connection between signifier and signified, but the link is purely conventional, mind-dependent – that's why there are many languages, in which we can express the same meaning, the same signified with different signifiers.

The brain is a physical system. It is undeniable that there are correlations between mental states / events / processes and brain states / events / processes, which are the object of investigation of cognitive neuroscience. Therefore, when we think then the structure of our brains and the processes occurring therein somehow reflect our thoughts and the meanings contained in them. In other words, there is some sort of physical-mental mapping. But, as we saw, such mappings are

necessarily indeterminate, i.e. whatever the brain analogues of series of bits in a computer or of sounds in a language, whether it is chains of connected neurons or something else, they can have no inherent correspondence to the elements of reality or meanings they represent. How can the arrangement of particles in relative motion interacting with each other through mutual forces that is our brain be naturally, and without the slightest help from an independent mind that decides a convention, representing a cat, time, space, a particular person, one's own self, freedom, numbers, justice, love, function, identity, shape, guilt, physics, World War 2, the universe, the philosophy of mind, or whatever else one can think of as an element of reality (or of a possible reality, in the case of imagination)? It is inconceivable that such a mapping could exist independently of a mind. Essentially, what the physicalist is forced to seek is a way to not only reduce consciousness to something physical, but also to reduce all of these meanings and concepts to something physical as well, to neural chemical structures.

Of course, one may argue that there is likely to be some sort of affinity between the physical signifier and the signified meaning. For example, in computers, our mapping of the bit sequence 00100110 to the integer 38 is not completely arbitrary but we exploit the affinity of the set of the two physical states "0" and "1" with the binary system of representing numbers. Similarly, in the case of the mind and brain, it could be the case for example, that when I am thinking of the concept of the number 1 then a particular neuron in my brain is firing, when I'm thinking of the concept of the number 2 then two neurons are firing in that same area and so on. But it does not follow logically that one neuron's firing is somehow inherently indicative of the concept of the number 1, any more than the bit sequence 00100110 by itself inherently means the number 38. After all, the concept of a neuron as a single entity is mind-dependent because a neuron is a complex object. And why that particular neuron when there are millions of neurons firing in my brain all the time? On top of that, even if it were the case that one neuron firing is somehow mysteriously indicative of the concept of the number 1, it would still not follow that the neuron's firing should induce in me the mental, first-person conception of the meaning of the number 1.

Alternatively, and equally inexplicably, it could be instead that when I think of the number 1 then a particular neuron is firing at a certain frequency, when I'm thinking of 2 then the same neuron is firing at twice that frequency and so on. Yet more alternative possibilities of physical arrangements and processes in the brain that have some affinity with the natural numbers can be thought of. If all I'm interested in is to produce some physical behaviour related to the natural numbers, e.g. I want to design a biological robot that touches its nose once if it sees one apple, twice if it sees two apples etc., then I could do that by utilising any of these neural designs, whether it is about the number of neurons firing, the firing frequency of a single neuron, or anything else that exhibits some numerosity. Producing something physical (the bodily behaviour) using something physical (the brain) as a computer is a "soft" problem, not a "hard" one. But producing something mental using physical machinery is not conceivable. Whatever the correlation between neural architecture & processes and mental thoughts, it does not follow from physical and logical principles but is primitive. In other words, strong supervenience of the mental upon the physical is inconceivable.

3.3. Weak supervenience

But how about weak supervenience? Do these primitive correlations cover every aspect of our mental lives, down to the last detail? Is the mental-physical mapping complete, i.e. are our mental states reflected perfectly in our brain states? If so, then deciphering the "physical language of thought" implemented in our brains will allow us to read someone's thoughts entirely by scanning his/her brain (assuming that the relevant technology has advanced enough to reveal all the necessary detail). However, the indeterminacy of this mapping, the lack of inherent meaning in the brain structures and processes, the richness of reality and of its mental representation, and the privateness of thoughts to the mind having them, suggest to me that the mapping is not complete and that perfect mind-reading in this way is not possible — hence, even weak supervenience does not hold. It should be noted that while the notions of integer numbers used here as an example can be, in a non-determinate way, mapped to brain structures and processes on the basis of the numerosity that the latter can exhibit, there are many other meanings, objects of thought, for which any physical signifier would have absolutely nothing in common with them. For example, most of the notions relevant to the philosophy of mind fall into this category; what neural pattern could have anything to do with, e.g., "philosophy" or "mind", or even "truth"? I therefore postulate that the patterns of neural activity in many cases only provide an indication of the general kind of mental activity that is taking place in the mind, without disclosing the full details.

If the mental-physical mapping is indeed incomplete, then this necessarily means that the mind is something separate from the body, and that it is the mind that does the actual thinking, and not the body. Of the neural processes that take place concurrently with the mental ones, the functionality of some may be completely explainable in terms of physics, such as those that are responsible for transmitting and distributing stimuli from the sensory organs to parts of the brain, and those that prepare and send signals from the brain to the rest of the body to govern its physical behaviour. Other processes may play the role of an intermediary between the mind and the former neural processes, in a mysterious way that is, as the present arguments suggest, inexplicable. Why, then, is there a need for the mental-physical correlations to exist, if the physical counterparts are not essentially necessary for the mental ones? And why do they have the particular form that they have, that is, why this particular physical language of thought and not some other, equally arbitrary one? I admit that these are difficult questions to answer, but, since the correlations are not part of a mechanism for the emergence of mentality, in my opinion they ultimately have to do with other kinds of causes that drive reality, besides the mechanistic ones that are the only ones recognised by physicalism; such drivers are purpose and teleology[9], which are compatible with a theistic worldview where the material world's sole purpose of existence is to provide a (likely transitory) pedagogical means for the maturation of minds. A detailed exposition of my views on this matter is beyond the scope of the present paper.

But even if the mind-body correlation is complete, the mere existence of correlations does not establish a causal relationship between the things correlated. For example, my shirt's sleeve is perfectly correlated with the motions of my arm, as long as I am wearing it, and if it is sewn onto my bed then I can't move my arm; but, if I take off my shirt, then my arm is free and the correlation ceases. A similar situation may obtain concerning the mind and the brain.

3.4. The physical inexplicability of mental phenomena

Physicalism employs various strategies, but they are all similar: aspects of the mind are identified with aspects of the

physical world with which they bear a superficial semblance or with which they are associated, such as structure, function, directedness, behaviour, information, causality. It seems that many people assume that the physical analogue (function, behaviour, directedness, structure etc.) *logically* entails (in a strong sense, like $a > b$ and $b > c$ entail that $a > c$) the emergence of the mental phenomenon to which it is likened — that there is a logical necessity that links the two. This is obviously a fallacy or wishful thinking. Nevertheless, this assumption is intuited by many people, including many scientists and philosophers (maybe especially them), as is made apparent, for example, by the large impact that Searle's Chinese room argument has had and the controversy it has stirred [27]. This argument merely demonstrates the (obvious) logical possibility that something, when acting according to instructions whose meaning is unintelligible to it (such as a computer executing a software code, or a man exchanging cards with inscriptions written in Chinese, a language he does not understand, according to an instruction booklet), succeeds in the Turing test, i.e. answers questions in a convincing manner such that a human cannot tell that it does not have real understanding of the questions or the answers. Surprisingly, many philosophers tried to refute the argument by proposing ideas that, in my opinion, are outright absurd, such as that the person in the Chinese room subconsciously understands Chinese, or that the room as a whole, including the human, the cards, the walls etc. understands Chinese.

Similarly, a behaviourist may think that if a robot of human shape is constructed and it is programmed to occasionally hold its belly or its head then it follows logically that during that time it feels pain. And a functionalist may think that if a toy car is fitted with a temperature sensor, a motor, and a simple processor that are linked so that if the sensor sends a signal to the processor that the temperature is too high then the processor sends a signal to the motor to move the car away from the source of heat, then the functionality of this heat-avoidance system logically entails that the car feels a burning pain when it touches something hot. Of course, behaviorism is easily refuted by thinking of an actor who holds his belly without feeling pain, or an ancient Spartan who feels a lot of pain without letting this reflect at all on his behaviour. Similarly, functionalism is refuted by thinking of someone whose brain is directly stimulated by electrodes at the right locations so that he feels that his hand is burning, whereas his hand actually has a normal temperature. But does one really have to resort to such examples in order to see the lack of linkage between the purported physical realiser (behaviour, function, etc.) and the corresponding mental phenomenon?

Suppose that it happened that once the Chinese room was set up it did exhibit understanding; and that the robot that held its belly did turn out to feel pain; and that the toy car did turn out to feel heat when close to a fire[10]. Would behaviourism or functionalism explain why this is so? Are these mental experiences logically deducible from behaviour or function? Is there any discernible logical link, or logical necessity between behaviour or function and mental experiences? I think obviously not. Therefore, either an *actual* physical explanation for the mental experiences must be found (possibly providing the actual missing link between function or behaviour and mental experience) or at least sought, or the correlation between the physical aspect (behaviour, function, directedness, information etc.) and the corresponding mental experience, assuming that it is universally holding, must be accepted as something primitive, inexplicable. The latter prospect, of course, is unpleasant to the physicalist, because it would imply that mentality is primitive and not reducible to anything physical, and therefore our reality exhibits some sort of dualism. But the physicalist also does not engage in a further quest to find the missing link between behaviour, function etc. and mentality. He/she pretends that the link is somehow obvious, or that behaviour, function etc. are identical with their purported mental correlates. This is

probably due to the realisation that this is the best he/she can do. If one wants to arrive at a destination but an impassable gap separates them from it, the best that they can do is walk up to the gap and proclaim that they have arrived. Extending the analogy made in Section 1 between physicalists and medieval alchemists, we can say that this strategy is similar to a medieval alchemist's attempting desperately to chemically transmutate lead into gold, and who, after many frustrating failures, discovers a chemical that paints the surface of the lead specimen yellowish, so that it now looks like gold from the outside; he therefore proclaims that he has succeeded — that's all there was to it.

Another problem with this physicalist strategy is that the physical aspects themselves which are considered identical to, or logically entailing, the mental phenomenon, i.e. behaviour, function, information, directedness, causality etc. are not entirely physical but necessarily have a subjective component. The physical reality comprises of countless purposeless and meaningless interactions between fundamental physical particles; minds recognise in macroscopic aggregates of such particles and processes structure, functionality, information, behaviour, directedness etc., but the way of doing so is subjective — different minds may see different things, depending on their focus of interest and their overall perspective. Obviously, if we consider any group of atoms or molecules interacting with each other, traveling with certain velocities, colliding with each other, pushing or pulling one another, then every single one of these particles can be seen to carry some information, can be construed to perform a certain function, to exhibit some behaviour, to give us hints about a nearby particle (directedness), to cause something to happen, etc. Hence, it seems that if physicalism considers consciousness to be fundamentally identical to, or realised by, information or function or behaviour or directedness or any of the rest, then it ends up being equivalent with some sort of panpsychism. In fact, it would assign consciousness not only to every single elemental particle, but to any conceivable group of such particles, such as a rock, or a pencil, or any part of a pencil, or half of the moon, etc. because functionality, information, causality and the rest can be found in any of these aggregates. Alternatively, one can contend that only some of these functionalities, behaviours etc. give rise to consciousness. But the placement of the dividing line seems arbitrary, or rather, it necessarily depends on the subjective perspective of a conscious observer, a mind. Hence in this case we would have a circular definition of consciousness.

3.5. A critique of modern physicalist theories of mental content

3.5.1. Causal theories

Let us examine one such physicalist strategy that links mental content to causation[25][28]. The central idea is that physical objects in one's environment cause things to happen in his/her brain and this constitutes his/her thinking of these objects, it constitutes the formation of the concepts of these objects in his/her mind. For example, suppose that a dog is in front of me and I am looking at it. As I am looking at it, I am also experiencing the concept of the dog in my mind; I understand that I am looking at a dog. Roughly speaking, these theories assert that the explanation to why I am understanding that a dog is in front of me is because a dog (and not, say, a tree) is actually in front of me and causes events in my brain by the light that is reflected on it and reaches my eyes.

It should be obvious that this explanation is nothing more than a superficial truism. It is a mere reformulation (usually expressed in philosophical jargon) of the very question it purports to answer. Undeniably, the fact that there is a dog in

front of me has something to do with my understanding that I am looking at a dog. This is obvious to everyone, whether a proponent of a causal theory or not. But the real question is why this physical causal chain that begins with light being reflected from the dog, continues with it reaching my eyes, and ends up with certain physical events in my brain, results in my mentally conceiving of a dog. The important question is not why I am conceiving a dog rather than a tree, but why I am conceiving anything at all, why and how inanimate physical events such as the reflection of light and the motion of ions in my brain result in qualitatively entirely different mental events. The answer to "why does a dog standing in front of me cause the mental event of me understanding that I am looking at a dog" cannot be "because this understanding is caused by the dog". Answering the question this way reveals a lack of understanding of the question.
Furthermore, the causal theory is tacitly circular in that it presumes that the entity that the concept of a dog represents is something objectively real, whereas if physicalism is true then a dog is just a composite object. It is a group of particles construed as an entity by a mind, for reasons that are subjective (see comment [7]) — not least of which is that the mind perceiving the dog projects itself onto it, regarding it also as a mind, as an embodied unified centre of consciousness. So, perceiving the dog as a unified entity presumes mentality; if the causal theory purports that our perception is wholly physically explainable then it should also provide a physical explanation to why we perceive the dog as an entity, but instead it naively takes the objectivity of the dog being an entity for granted. In other words, the causal theories presuppose the subjective structure of the world that we perceive, which they are supposed to explain.

But perhaps the aim of causal theories is not to explain the emergence of mentality from physics, but merely to discover the correlations between physical and mental events, accepting them as primitive. But even if this were the case (which is not, because the aim is to naturalise, i.e. "physicalise" mentality), the physical correlate that these theories propose, which is a physical causal link between the element of external reality that a person perceives and that person's brain, seems incorrect and inadequate. Rather, the physical correlates should be sought among the physical structures and events *inside* the brain [26]. Of course, these could, in turn, be caused by external stimuli, but this is not necessary, and even then the external environment would only be causing the mental events indirectly, by causing brain events (which in turn mysteriously cause mental events). To make this more clear, consider again the example of me looking at a dog. There is a causal chain which consists of many links: light is reflected on the dog's skin and fur; some of it reaches my eyes and stimulates photoreceptor cells in my retinas; these produce neural impulses mirroring the light signals they receive, which travel to other parts of my brain; the neural machinery there processes and extracts various kinds of information from the incoming signals; finally, somewhere towards the end of this causal chain there is a missing, mysterious, inexplicable link: the link that connects some physical event(s) in my brain with the mental event of my perceiving a dog. That the first link of this chain, the light being reflected from the dog's skin or fur, is not directly implicated in producing my perception of a dog, is very plausible since I would get the same perception even if I were looking at a picture of a dog or a movie of a dog in my computer screen, without any actual dog present — even if I had never seen an actual dog in my life. Furthermore, I can think about a dog when one is not present. Also, various unreal perceptions are generated in my mind if I am wearing 3D glasses or virtual reality headsets; if I am in the metaverse, then I perceive of a whole world that does not really exist. The brain itself has mechanisms that modify our sensory input, and produces visual illusions such as afterimages. The scientific experience is that what we perceive is determined by the physical processes inside our brains, the consensus being that if we reproduced exactly, say by the use of electrodes or magnetic stimulation etc., the processes that occur in

one's brain onto another person's brain then the second person's mental state would be exactly the same as the first person's, even if their actual environments were entirely different. This means that if my brain is artificially stimulated in precisely the way that a brain is stimulated in the presence of a dog then I will perceive a dog even if no dog is present, and even if I have never before experienced perceiving a dog. Furthermore, and more importantly in my opinion, many mental concepts are not about physical things in our environment and hence no candidate physical correlate in the environment exists anyway. How can causal theories explain my conceiving of immaterial things, such as the concepts of mathematics or mental phenomena or persons or of abstract concepts such as value, freedom, love, justice, safety, truth, reality etc.? Hence to discover the physical-mental correlations we should focus on the correlation between mental events and physical events inside the brain, while physical events in the external environment (which causal theories regard as of prime importance) are, ultimately, irrelevant.

I think that these arguments suffice to show the absurdity of the basic premise of causal theories of mental content. Causality is ubiquitous in the physical world, which consists of countless fundamental particles in constant interaction with each other, pushing and pulling each other around. All of these interactions can be viewed as causing things to happen. Does this, according to causal theories of mental content, result in mental events? If so, then these theories seem congenial to panpsychism rather than physicalism. The dog is surrounded by air, the molecules of which are constantly bumping against the dog's body, in perfect correlation with the dog's shape; does this mean that the surrounding air understands that there is a dog there? Does the ground understand that there is a dog there because the dog causes a dog-shaped shade on it, and exerts a dog-shaped pattern of pressure on it due to gravity? Why is there a need for our brains to have the complex structure that they have, and what makes them special, from a physical point of view, compared to other physical objects such as air or the ground that can also receive influences from their environment but do not seem to understand anything or be conscious? These are the important questions that arise if one considers the possibility of a physical foundation of mentality, but proponents of causal theories seem oblivious to them.

3.5.2. Causal theories supplemented

Some causal theorists recognise that, since causal interactions are ubiquitous in nature, causation is too wide and general of a concept to be a realistic candidate for a physical substrate of mentality (unless one is a panpsychist). Hence they propose causal theory versions where it is not just any physical causation that gives rise to mental content, but only causation that either has some purpose (a teleological view) [29], or some function [23], or transmits some information [30][31] — actually, it seems to me that the differences between these variants are very minor; essentially they are the same theory: mental content is produced by causation whose *purpose* is to *function* as an *informer*. They place emphasis on minor problems of causal theories of mental content, especially on how misrepresentation can arise (e.g. how the theory can accommodate for the possibility that I think that I am looking at a dog when in reality what is in front of me is a fox), which is a soft problem, and seem blind to the hard problem of the impassable gap between physical causation and mentality — they strain out the gnat but swallow the camel.
Furthermore, they seem oblivious to the fact that all of the concepts employed — purpose, function and information — are subjective and meaningful only with respect to a mind. Such concepts belong to the "intentional" and "design" stances of

Dennett [24], as mentioned in comment 9, whereas the only relevant concept from a “physical” stance is causation. Hence, in physicalist reality, the causation associated with what these theories call "purpose", "function" and "information" is indistinguishable from any other physical causation that these theories have filtered out as possible generators of mental content, and the dividing lines are completely subjective.

For example, an earth pillar is a naturally occurring geological phenomenon consisting of an upstanding column of soil on top of which sits a stone that protects it from erosion. The earth pillar took thousands of years to form, as the surrounding unprotected soil eroded. In intentional or design stance language one could say that the function or purpose of the stone is to protect the underlying soil from erosion, but clearly this is only a figure of speech as the stone actually has no purpose or function but happened to be there by chance. In physical stance language we could simply say that the rock on top of the column causes the underlying soil to be inaccessible to the eroding elements of the environment (rain, for the most part). The reason why one could view this causal effect of the stone as a function is that subjectively a pillar appears to a mind as a peculiar formation that stands out from the rest of the natural surroundings, and the mind assigns some special importance to it. However, objectively it has no special importance and in fact is not even an objective entity, being simply a composite, mind-defined object (Comment 7).

If physicalism is true, then the exact same holds for the physical systems that we call biological organisms. From an intentional or design stance, which are subjective, we say that the purpose/function of the heart is to pump blood to the tissues to nourish them; but from the objective physical stance the heart causes a pressure difference that moves blood along the arteries from where oxygen and other chemicals diffuse into the tissues due to concentration gradients. Objectively speaking, the heart has no purpose or function; it developed over millions of years of random purposeless mutations that occurred due to chance. Just like the rock on top of the earth pillar does not have the purpose of protecting the underlying soil but merely happens to cause it to be protected and hence the pillar is preserved for thousands of years by chance, so the heart does not have the purpose of nourishing the body but merely happens to cause it to be nourished, and hence the body endures long enough to reproduce itself (through other mechanisms that also occurred by chance) and copies of the organism exist for thousands of years. Maybe a factor that confuses people is that, whereas the transportation of the stone to the location where the pillar subsequently formed seems to be a one-step chance event, the formation of a heart in organisms occurred in a multi-step process that spanned millions of years. However, each of these steps is exactly like the chance event that caused the stone to move there; in fact the motion of the stone there also occurred in a series of steps that lasted millions of years.

Or, consider again the example of Section 3.4 of the toy car fitted with a processor, a motor, and a heat sensor, that is programmed to avoid heat. In this case, these components do have a purpose and a function in the sense that the car was designed this way by a mind (a human person) with these intentions; it is relative to that mind that the components have their purpose and function. But if we consider a simple biological organism of similar “design” but which is made of biological technology, which was not designed by anyone but evolved according to natural selection over millions of years, then none of its components have any purpose or function, if viewed objectively and mind-independently. What happened during these millions of years is that many mutations occurred by chance as the organism's ancestors were reproducing; many of these mutations resulted in offspring that did not themselves reproduce because they were

destroyed by heat before they had the chance to do so. Such mutations may have caused the absence of heat sensors, or may have caused neural structures that caused the organisms not to move away from the sources of heat, or even to move towards them and get destroyed. On the contrary, those of its ancestors in which mutations occurred that caused them to have heat sensors and to have such neural circuits that cause them to move away from the sources of heat had smaller probability of getting destroyed and thus larger probability of reproducing and passing on these mutations. There is nothing special about the mutation of the extant organism compared to the mutations of the branches that perished, except that the former causes increased probability of reproduction while the latter cause lower such probabilities. That's all there is to it. It did not occur by design but it is a mere contingency. Evolution by natural selection is a contingent purely mechanistic procedure with no purpose. The fact that it reaches equilibrium states may seem like a purpose to many people, including the aforementioned causal theory proponents, but this is only due to our natural inclination to think in intentional-stance terms because this is the natural stance for minds.

The intentional stance inclination is manifested also in the fact that these causal theories assume organisms to be objective entities, natural kinds, whereas if physicalism is true then what we call biological organisms is nothing more that composite objects, which have only a subjective existence, relative to mental observers (see Comment [7]). Our natural intentional stance inclination often causes us to project our own sense of ourselves as unified centres of consciousness onto objects of our environment, personifying them; some of them may be actual persons but some may not, e.g. an individual cell or even more complex organisms. Since we have a sense of our own value as living existences, we regard all life as valuable. This leads subconsciously to the idea that not only do organisms themselves want to preserve their own life but it is also a universal goal of nature to preserve life, that teleology concerning life is embedded in the fabric of nature. This idea seems to lie at the foundations of causal theories based on evolution, purpose and function, but the only way to construe the biological evolutionary process as having purpose, and the components of the organism as having a function conducive to the attainment of this purpose, is to accept that the physical world is governed by some higher principle that has intentions and purposes and is therefore mind-like. Therefore these theories instead of explaining mentality in terms of physics, actually suppose that physical aspects of the world are governed by a fundamental mentality that lies in the foundations of reality. They deviate from the spirit of physicalism and have an affinity to religious views.

3.5.3. Informational theories

There is a relatively recent class of theories, called "scientifically guided informational theories" in[31] (see references therein), which avoid intentional notions and use mathematical language — probability theory in particular. These theories claim that the representation of a physical entity is that structure in our brains which is most likely to become activated when we perceive that particular entity; and conversely, that when that neural structure is activated in our brain, it is most probable that we are perceiving that specific entity (of which the neural structure is a representation) rather than any other entity. These theories attempt to express in a formal manner the intuitive empirical practices of cognitive scientists. For example, suppose that a scientist monitors the activity of the brains of several human subjects as they are shown various things, and notices that when shown a dog it is usually a particular neural network — let us call it 'DOG' in hindsight — that becomes activated. Conversely, the scientist notices that when 'DOG' happens to be active it is usually the case that the

subjects are currently looking at a dog. Hence, the scientist draws the conclusion that the neural network 'DOG' represents a dog. The informational theories make a law out of this empirical observation.

It is tempting to regard such theories as doing a better job naturalising mental content than the causal theories previously mentioned, since they avoid references to subjective concepts such as purpose, function, and even information in the subjective sense, despite their name. However, with a closer look it should not be difficult to see that they are devoid of theoretical significance and of explanatory power. As sketched in the previous paragraphs, the process of perceiving a dog in our scientist's experiments consists of a causal chain of physical events whose start we could (arbitrarily) place at the reflection of light on the dog's fur, and whose end (as far as perception is concerned; the causal chain may contain further links that produce behaviour) is the activation of the neural network 'DOG' which inexplicably causes in our minds the mental event of perceiving a dog[11]. Intermediate physical processes perhaps have the role of decomposing the incoming signal into components pertaining to colour, shades, aspects of geometry, pattern of motion etc., the relative strength of which eventually determines, through the mechanics of processing circuitry, that it is 'DOG' that eventually becomes activated rather than, say, 'TREE'. Of course, the correlation between dog and 'DOG' is unlikely to be perfect, and occasionally mistakes or misrepresentations will occur; perhaps I am actually looking at a fox but it is dark and the incoming signal lacks sufficient strength or detail, or even I am looking at a dog-shaped cloud and playing with my imagination. Since there are many kinds of dogs, and a dog may be in different states, viewed from different angles, under different lighting conditions etc. 'DOG' should become activated under a range of incoming signals, not only for a unique signal; this increases the possibility of error. Hence it is expected that when one is looking at a dog then usually, but not always, 'DOG' will fire causing the mental concept of a dog to arise in one's mind; and conversely, when 'DOG' fires in one's brain it will usually be the case, but not always, that they are looking at a dog (this latter statement is less likely than the former, because one could be merely thinking of a dog or remembering one, or watching a cartoon with dogs, or reading or being told about a dog etc. without any actual dog being in front of his/her eyes). It is this correlation between dog and 'DOG', for instance, that these informational theories attempt to quantify probabilistically, proposing probability as the criterion for matching objects in the external environment such as a dog to representative brain structures such as 'DOG'.

But it should be obvious that these informational theories do not explain anything. Rather, they filter out all the substantial details of the aforementioned causal account and merely provide a rule of thumb for identifying the neural structures associated with the mental representation of material objects that are detectable by our senses. This rule of thumb is not even always correct, as noted in [31], and hence it cannot hold the status of a law. That it will *usually* be correct makes sense, if one considers the more detailed causal mechanics outlined above; but then, these informational theories are explained by these mechanics, whereas they themselves explain nothing. They overlook the core of the issue, the hard problem, of why and how something material such as 'DOG' causes the mental experience of thinking of a dog in someone's mind. As good as the correlation between a dog and 'DOG' is, that between the dog and the mass of air that surrounds it is better. Why is it that this mass of air does not mentally understand that there is a dog there? The informational theories are silent on this, in a way that suggests that they do not even see the problem.

Furthermore, like causal theories they are also short-sighted in that they consider only the perception of material objects, and overlook the conception of theoretical, abstract and immaterial notions and ideas. Of course, informational theories

could easily be extended to account for these by correlating them to the neural structures that are most active when the subject is thinking about these concepts; for example, if whenever subjects are asked to judge whether something is right or wrong the neural network 'JUSTICE' is observed to fire in their brains then it is reasonable to assume that it is 'JUSTICE' that is inexplicably tied to the mental conception of fairness. However, this is consistent with an empirical rule for discovering the mind-body correlates and not with a theory that purports to naturalise mental content.

4. Persons: external symmetry and internal asymmetry

Although their depth is generally underestimated, the aforementioned hard problems of consciousness and mental understanding are recognised in the philosophy of mind. On the contrary, the problem that we are about to turn our attention to, that of persons, is hardly ever brought up in contemporary discussions, even though, in my opinion, it is the hardest and deepest one. Introspection reveals to anyone that he/she is a centre of conscious existence, of first person-perspective, characterised by simplicity (non-compositeness) and uniqueness (non-duplicability), properties that are not consonant with the physical realm and which are not compatible with a physicalist view of persons as physical objects. The problem is multifaceted, and we will look at it from various aspects in an effort to gain as deep as possible an understanding.

4.1. The pairing problem

Whether considered as a literal fact (by dualists) or as an illusion produced by biochemical processes (by physicalists), we can say that each of us is a centre of existence, a mind that thinks, feels, senses, reasons, etc. which is interwoven with a particular body: *I* see through *my* eyes, *I* can raise *my* hand, *I* feel pain if *my* foot steps on a nail, *I* loose *my* intellectual powers when *my* brain suffers from Alzheimer's disease etc. I don't have this special connection to any other body in the world, nor does this particular body of mine have such a special connection to any other person in the world. A natural question to ask is how each person is paired to a particular body. Why am I paired to this body and not to some other? What determines this? Trying to answer this question reveals another hard problem for physicalism: it is impossible for there to be a physical mechanism to determine this pairing.

Interestingly, a version of the "pairing problem" has been put forth by physicalists as a hard problem for dualism (e.g.[32]; see [33][34] for refutations). The argument revolves around causation, and contends that, since souls are not located in space, they could not cause physical events in the body as there can be no spatial connection between the soul and the body, something that is normally required in the causal relationships we observe in the physical world. The argument does not carry much weight and, like other physicalistic arguments against dualism, assumes its conclusion, since the "problem" is ultimately just that dualism is not compatible with physicalism. For example, an explanation that God has determined the pairing of each one of us to his/her body on the basis of which body would be most beneficial for him/her would be outright rejected, but a careful and honest introspection may convince the physicalist that the reason for the rejection is just that the explanation is not physicalistic, is not based on of physical principles.

A more substantial and meaningful pairing problem is the question that each of us can ask: what is the cause of my pairing to this particular body that I am paired to? Speaking as if physicalism is true, out of the billions of bodies currently alive on earth, why is it that mine, and only mine, gives rise to me? Why is it that I am experiencing life through this particular body and not, say, through a particular female body somewhere in China, or a particular 60-year old male body in Brazil, or a particular body of a child in South Africa, or the body of my brother, or that of my mother? Furthermore, it seems to be the case that once a particular material composite, a body — my body — has given rise to me through intricate physical, chemical, biological structures and interactions, thenceforth "the seat is taken" and no other new body that is formed / conceived, even if thousands of miles away (so that it does not interact in any way with my current body, it is "unaware" of its existence), is allowed to also give rise to me: I cannot be paired to two bodies at the same time; I would have to be two persons at the same time, since each body has its own memories, its own stream of perceptual input etc. It seems very unlikely that such a prohibition could be explained physically.

A physicalist may try to dismiss these questions as arising from a false premise that I am a separate entity from my body. But if we concede for the moment that my perception of myself as a separate entity associated with my body is just an illusion and in fact I am identical with my body, nevertheless that illusion still deserves explanation. If everything ultimately comes down to physics, then this illusion should also be physically explainable: there should be a physical explanation for why this specific first-person perspective phenomenon I perceive as myself is generated by this particular body and not some other. Sure, every other body will also presumably generate such a first-person phenomenon, but it will be *another* one, not mine, even if exactly similar as viewed from the perspective of the person paired to that other body. If reality is entirely physical, and therefore all aspects of it can be explained in terms of physical principles, then the fact that I perceive myself as bound to this particular body, whether my perception is an illusion or not, should be explainable in physical terms.

To show that the cause of the particular mind-body pairings cannot be a physical mechanism, and that selves are not reducible to bodies, consider the following fictitious experiment: suppose that, many years into the future, technology has advanced such that there are 3D printers that can print any arrangement of molecules we desire, even a human body, with the molecules at exactly the right places such that the body is instantly functioning and alive. Using this printer, we make an exact copy of a living person. Both persons have exactly the same bodies, meaning also the same brains, which arguably means that they would have the same memories (the new person would mistakenly think that he/she is the original person), they would have the same intellectual capacities, they would like the same music and food, etc. Everything that can be mapped to a physical structure in their bodies would be the same. But there would be a very important difference: they would not be the same person. Imagine that you are one of these persons; say, you are the original person, and a physical duplicate of you has been created. If someone grasps the duplicate body's foot, would you feel it? If someone places something in front of the duplicate body's eyes, would you see it? From your own, first-person perspective, clearly there is a huge difference between the two bodies: you are, or "inhabit" (depending on whether physicalism or dualism is true), only one of the bodies — you are paired to only one of them. But, if it is the intricate biological machinery inside your physical brain that gives rise to you, and the duplicate body's brain's machinery is exactly the same as yours, then the new brain should also give rise to you. However, it does not; it gives rise to another person[12].

If physicalism is true, then the pairing of bodies to persons should be determined by a physical structure / mechanism, and there should be a physical explanation for why you are mapped to the original body and the other person to the new body. But obviously there cannot be such an explanation, since the physical arrangements of the two bodies are identical. The structure of the body cannot be mapped onto the person, it doesn't tell us who is who.

One could argue that although the structure of the two brains is the same, still they are two separate pieces of matter, hence the two persons. But this objection is not very convincing, since the crucial aspect of the brain seems to be its structure / function rather than the particular molecules it consists of. The matter that constitutes our brains changes all the time (in fact, some of it may have previously belonged to other people's brains), yet we remain the same persons. To remove any possibility of doubt, let us extend the previous thought experiment in a way analogous to the thought experiment about the ship of Theseus. Suppose an oxygen molecule from your brain is swapped with an oxygen molecule in your duplicate's brain. Arguably, this will not have any impact on you whatsoever, since your brain structure has remained the same, and our brains change molecules all the time, yet we remain the same persons. Repeating slowly this procedure, we can end up swapping all molecules between the two bodies, so that you now possess the body originally owned by your duplicate, and he/she possesses your original body[13].

The ability to make an exact copy of a body will not be available anytime soon, but this does not take anything away from the power of the argument. Besides, we can ask the same question for monozygotic twins, whose bodies are, of course, not exactly the same but are very similar, and were almost exactly the same during their early stages of development. But furthermore, the requirements of exact similarity can be relaxed since, e.g., my body changes every day but I am still the same person, myself, mapped to the same continuously-changing body. Therefore, there is a huge set of bodily configurations, from when I was a baby until I grow old and die, that map to the same person, me. So, we have situations where different bodies pair to the same person (e.g. the body I had when I was a child and my current body both map to me), and other situations where identical bodies pair to different persons (e.g. in the above thought experiment, or the monozygotic twins case if we allow for some small differences between the bodies). Therefore, the physics of the body cannot account for the totality of mental phenomena, and in fact it can not account for the most important aspects of them.

It is noted that the present arguments pose a hard problem for physicalism even if the latter is defined in terms of supervenience. If two identical bodies produce different persons, e.g. you and another person as in the aforementioned thought experiment, then the person cannot be supervenient on the body. To make this more convincing, add on top of that the Theseus-ship-type molecule swapping experiment, so that originally you were paired to your own body and the other person to the duplicate body, and at the end of the experiment you are paired to the duplicate body and the other person to your original body. Of course, from a third-person point of view, you and the other person are completely similar, completely symmetric. But from your (or the other person's) point of view, there is a remarkable difference between yourself and the other person, a striking asymmetry. Both the third-person symmetry and the first-person asymmetry (which will be discussed further in the next paragraph) are part of reality, so physicalism should account also for the asymmetry, but it cannot.

4.2. Third-person symmetry and first-person asymmetry among persons

The previous thoughts about the pairing problem can serve as a warm-up to proceed to deeper thoughts and arguments about the self. If we dig deeper into the nature of the self, then not only can it be shown that the pairing problem is a hard one for physicalism without resorting to thought experiments involving identical bodies, but also that any theory that claims that the self is explainable by factors external to it only, be they physical or mental, is problematic.

If we introspectively examine our own mental selves, we see (I assume, since my own self is the only self I have direct access to) that we are all different in many peripheral respects, such as our memories, our beliefs, our tendencies, our likes and dislikes, the way we think, our mental and emotional capacities, even maybe the way we perceive — one person's blue may be another person's red, or one person may lack the sense of vision altogether. Furthermore, the current mental state of each of us is different: one is reading this paper and contemplating about it, another is driving her car to work and simultaneously listening to music, another is asleep and dreaming, another is trying to figure out ways for himself and his family to survive in a war-stricken area, another is enjoying the company of people she loves, feeling grateful and content about her life, while another may be struggling with thoughts about suicide, unable to find meaning in life. However, at their core, all persons are fundamentally similar: each is a centre of existence, of first-person perspective, of consciousness, each is his/her own self, a person, an ego. We are equally alive. In this respect, we are all equal, we are all the same. No one is any less his/her own self than any other. Of course, this is just a hypothesis from someone who has no direct access to anyone other than his own self, but it is a very plausible one (and intuitive one, since it is the basis for empathy, putting oneself in another's place), since being an ego seems to be an all-or-nothing thing; it is not possible to imagine different types or degrees to it. I will refer to this fundamental similarity between persons as *symmetry*.

Again, I emphasise that this does not refer to the peripheral characteristics which may be different from person to person, but to the core quality of being a person, a centre of existence, a centre of consciousness. This quality is associated with the question of *who* a person is, while the peripheral qualities tell us about a person's mental state and about his/her character, and could potentially be swapped between persons; today I am happy and another person is sad, tomorrow I could be the one who is sad while the other person is happy — but I still remain myself, and the other person remains him/herself irrespective of the peripheral mental changes we experience and undergo.

Although objectively, from a third observer's point of view, all persons are symmetrical i.e. exactly similar, from a subjective, first-person point of view there is a fundamental difference between one's own self and all others. From each person's own perspective, among all existing or possible persons only one is singularly different from all others: his[14] own self; he is his own self and not any other, he directly experiences his own self and none other. Of course, intuition and reason lead a person to believe that the situation concerning all other persons is symmetric to his own; that just like for him, for any other person, from their own point of view, the singularity concerns their own self compared to all others.

Let us explore this singularity a bit further. Between me and the rest of reality there is a discontinuity. I cannot continuously change into someone else or something else (of course, my peripheral qualities can). Being myself is an all or nothing thing. Similarly, there are no persons or things that are more "me" than others, e.g. one other person is 25% me, but another is 50% me. All of them are 0% me and I am the only one who is 100% me (a person peripherally identical to me, such as my bodily duplicate of the thought experiment of the previous section, is still a completely different person from

me, inaccessible to me. He is 0% me). Likewise, there is a barrier between me and all material things; they are all equally foreign to me. There are no oxygen molecules that are more 'me' than others. By this I mean that the particular first-person point of view that is characteristic of me cannot be found in any oxygen molecule, whether inside or outside of my body; they are all equally unconscious (and even if they were conscious, their consciousness would be completely disjoint from mine), none of them being more similar to me than the others. Nor am I more special to them than any other person is, since all persons are exactly the same to all entities other than themselves. The same holds for structures. The neural circuitry in my brain, although associated with me, is not more me than any other circuit; in fact, as argued in the previous section, the exact same circuitry in another brain would be associated with another person.

With these considerations, let us revisit the mind-body pairing problem. What is so special about my own body such that it gives rise to me and not to someone else? By "special" would be meant a characteristic of my body that ties better to me (stripped from all my peripheral qualities[15]) than to any other person. But since all persons are symmetrical, no such characteristic can exist: any special feature of a body would tie equally well with all persons, since all persons are symmetric from an outside point of view. Consider two persons, say myself and my brother. If I prefer classical music and my brother prefers folk music then sure, this could be attributable to the relevant structures in our brains being different. But this does not determine that I should be paired to my current body and my brother to his. Why could I not instead be paired to my brother's body and he to mine, in which case our musical tastes would be swapped along with our brains (I would prefer folk music and he would prefer classical music)? Arguably, all our peripheral characteristics, such as our memories, our current thoughts, the current inputs from our senses, our likes and dislikes, etc., can be mapped to structures and processes in our bodies. But what about the most important characteristic: who is who. Why is my body paired to me, and my brother's body paired to him? Since all persons are exactly similar in this respect, no particular physical feature of each body can be considered responsible for giving rise to that particular person that is paired to it instead of any other (existing or possible) person. The pairing cannot have a physical explanation.

What is so special about your body in relation to you, such that it determines that it is you that it gives rise to and not someone else? A contemplation will reveal that there can be nothing special about it. No DNA sequence, neural architecture, atomic composition, shape, weight etc. can have any a priori special connection to you compared to any other person; all such features are equally neutral towards all persons. The specialness of your body towards you does not follow and cannot follow from any of its physical characteristics; it is only special to you *because you happen to have this body*; it is an a posteriori specialness. There is no physical cause that can determine that this body shall be mapped to you and not to someone else.

Similarly, each one of us can make the thought experiment of going back in time until before he/she existed, and wonder why, when his/her body was formed, it was him/her that came to being, paired to that body. Would there be any difference if it was someone else that came to being instead of me, when my body was formed? It would be a person exactly the same as me in every respect, except one: it would not be me but someone else. But this crucial difference is only a difference from the perspective of the persons affected, me and the other person. For the rest of the universe, there is no difference at all. There is an infinity of other possible persons, selves, that could have come into being instead of me, and they are all indistinguishable from me from the outside, to a third observer, including the inanimate physical world.

These thoughts highlight the impossibility of there existing any a priori special connection between a person/mind and elements of the universe outside of it. As such elements we mostly considered material bodies and physical properties, because that is the focus of physicalism, the prevailing view about the mind and persons. However, it is evident that such elements are not restricted to the material realm but could be anything, even other minds/persons. For example, in exactly the same way it is impossible for there to be any *a priori* special connection between a person and his/her parents compared to any other third person. All other persons, including my parents, siblings etc., are completely symmetric from my point of view and none of them has a first-person perspective that is closer to mine than that of any other person. For each of us, our own first-person perspective is something unique, and those of others are equally 'third', foreign and inaccessible to us. Hence, there is nothing that can a priori explain why my parents would beget me specifically among all possible persons. The same holds for any purported micro-consciousnesses associated with the fundamental particles of matter that panpsychism views as constituting a macroscopic person such as you and me. This will be discussed further in the next section.

Summarising, from the point of view of a particular person, his own self is a singularity among all of the rest of the universe. However, from the point of view of the rest of the universe, that person is exactly the same as all other persons. This raises the question: how can a person's existence be explainable with reference to factors only external to it? If the singularity, particularity and uniqueness of a person can be found only within that person itself and nowhere else in the universe, then where does it come from? We will consider this question in Sec. 4.4., but first let us consider the question of whether a person is a composite or a simple entity.

4.3. The composition problem

With the aid of the previous arguments, let us examine whether the person/ego/self is a composite entity, analysable into constituents. Several worldviews or schools of thought regard it as such. Among these is physicalism, which regards it as ultimately a macroscopic manifestation of a combination of physical processes involving a huge number of fundamental physical particles and their interactions. Another such view is panpsychism [17], which regards a person as an aggregate of a huge number of elemental consciousnesses (or proto-consciousnesses), the latter being an intrinsic property (or the intrinsic nature) of fundamental physical particles. Still another example of such a theory, encountered in the philosophy of religion, is traducianism, according to which a new soul is produced from individual contributions from its parents, in a similar fashion to its body. Each of these theories may come in many different flavours, but they all assume that persons are not fundamental but are phenomena that can be explained by reference to their constitutive fundamental parts and how they are arranged and interact with one another. However, it will be argued here with reference to the symmetry argument, that the core of being a person, of being someone, *who* someone is, cannot be explained by reference to any constitutive elements. Persons are simple and fundamental entities.

I, as a person that is symmetric with all other persons, do not bear any special relationship to my alleged constituents, be they physical particles or elemental consciousnesses or the structure of my brain or parts of my parents' souls etc., compared to the corresponding alleged constituents of other persons, that could explain why (or determine that) they should give rise to me specifically and not to some other person. If I could hypothetically travel back in time to before I

existed, and examined all possible combinations of material particles, elemental consciousnesses, or pairs of parents, it would not be possible to determine beforehand that one of them, the particular one that eventually was brought about by my parents' union, would give rise to me. Why that particular combination and not some other one? And conversely, why did this particular combination, induced by my parents, give rise to me and not to some other person, if all persons, including me, are exactly similar from the constituents' perspective? Do the molecules that make up my body have anything more in common with me than with any other person? No, because all persons are the same. Does the structure of my brain have some distinguishing feature compared to all other existing or possible brain structures that relates better to me than to any other person? No, because I am "I" in exactly the same way as anyone else is their own selves. Do I have something more in common with my parents than with any other person? No, since all persons, including my parents, are equally third, inaccessible persons to me, as I am to them. The same holds for my relationship to any hypothetical elemental consciousnesses: they are all the same to me — none of them is more "me" than any other — and I am the same to them as any other person.

Any combination must be a priori equally neutral towards any person, due to the symmetry between persons. Two persons are exactly similar from any point of view outside of themselves, including the perspective of any alleged constituents; so, there is nothing to grasp onto to make the pairing between combination and person determinable. There is nothing to grasp onto to allow that the formation of a certain combination is the cause of the emergence of that particular person. But, the coming into existence of a person coincides with its pairing to a body; the latter is indeed a composite entity, and its formation requires a process. Hence, it is intuitively assumed by many that that person's coming into being has been determined by the particular process and combination that led to the formation of the body. But this is not possible, because all possible persons are exactly similar to each other and no procedure or combination can determine a particular person among a pool of infinite identical possibilities.

Consider another thought experiment: Suppose I have the skills and resources to compose a human. I can design his/her body down to the most minute detail, to the last molecule. Overlooking the hard problems of Sections 2 and 3, let us assume that I can thus set his/her memory, intellectual ability, desires, whatever — all the aforementioned "peripheral" qualities[16]. But how can I set *who* he/she will be, i.e. which ego/self will inhabit or emerge in that body? This is completely out of my control, despite all aspects of the body being directly under my control. All possible egos, infinitely many of them, are exactly similar, completely symmetric; they have no objective difference, but only the (crucially important) *subjective* difference that for each one of them their own self is singularly different from all others in that they experience only their own lives. So, there is no way for me to select a person among the infinite possibilities, given that they are all the same to me; from my perspective, they don't have anything different between them that I can associate with a particular bodily characteristic. However I design that body, e.g. whatever DNA I design, or whatever neuron connectivity I design in the brain, or whichever molecules I put at any location, all possible egos are still equally associatable to that body. Therefore, if the person that arises due to my composing his/her body asks me: "How come this body gave rise to *me* specifically and not to someone else?", or "why did I not emerge from any other body in the world?", what could I answer him/her? There is nothing that I could answer, despite my determining everything concerning his/her body.

The same problem also occurs if I were free to choose the elemental (proto)consciousnesses of the physical particles composing the new person's body, if pan(proto)psychism is true, or the person's parents (traducianism). Obviously, I could

not answer that person's existence question either in terms of elemental (proto)consciousnesses or of parents, as all (proto)consciousnesses and all parents bear the same, third-person relationship to all potential persons equally; they are equally outsiders to both the person actually brought into existence and to all other fundamentally identical potential persons that were not created.

Let me insist on this point a little further, by changing role and taking the place of the created person instead of the engineer. Suppose I am the person just created. If I ask my engineer "why am I good in trigonometry?", he might reply "because I designed the neural circuits in that part of your brain in such and such a way". If I ask him "why do I have a hot temper?", he might reply "because the design of your body is such that it produces an excess of that substance, which causes those neurons to be triggered more easily". But if I tell him: "That's all nice, but there is one pressing question that is burning me above all others: why *me*? Why did it have to be me that was created?", then he could not answer me, because he did not determine this, he could not have.

Note that this problem has two sides, which compound its hardness. On one hand, from the perspective of any candidate constituents, all persons are exactly similar so there can be no reason why a certain combination will produce one person rather than another. On the other hand, the difference between persons is subjective, observed only from within each person. Each person's existence is characterised by privateness, that only he/she has direct access to his/her own self. This is what makes a person and identifies it compared to all others. And it is this that we are trying to explain. But this can be found nowhere else in the universe except within that person him/herself. So how can it arise from a combination of things that do not have it? Whether I consist of physical particles or elemental consciousnesses, these are all foreign to me, as foreign as the constituents of any other person. None of these outside candidate constituents, whether mental or physical, is any more "me" than any other, and in fact, all of them are 0% me whereas only I am 100% me, as mentioned previously. So, how could the singularity of "me" arise from any combination of things that are completely disjoint from me?

This problem is related to the "combination problem" of panpsychism, the most influential formulation of which was given by William James [35]. There are several variants of this problem[20], but the one closest to the present arguments is essentially the one put forth by James [35], called the subject-summing problem (here the term "subject" is used synonymously to what we have called person or self in the present paper (macro-subject), or to the elemental consciousnesses of panpsychism (micro-subjects)): the aggregation of a number of subjects does not necessitate the formation of a new subject. In fact, the present arguments show that not only do they not necessitate it, but it seems impossible that a subject is the aggregate of other subjects. These constituent subjects have nothing more in common with their alleged aggregate subject than with any other subject. To each constituent subject all other subjects, including the alleged aggregate one, are equally inaccessible, privateness being a distinctive characteristic of subjects.

To the present author thoughts such as these suggest very strongly that persons / egos are at their core not composite but simple substances. In fact, I would go as far as to say that an ego is the quintessential simple substance.

4.4. The creation problem

The reader who has followed the discussion thus far may have anticipated the next problem, which arises naturally, is even harder than those previously mentioned, and goes beyond physicalism and panpsychism. Let us grant that a self is a simple entity; how then is it brought into existence? A deeper contemplation about this question, considering especially that the self is characterised by privateness — that only an ego itself has direct access to its own self, while the rest of reality is isolated from its inner world — makes it seem impossible that only factors or agents external to an ego can bring about its existence. To appreciate why this seems impossible, it must be taken into account that "bringing about the existence of an ego" includes bringing about the most crucial part of an ego: *who* that ego will be, which is something completely private.

Consider again a thought experiment where now instead of being an "engineer" of persons you are a creator of persons: you have been bestowed with the mysterious power to create any person you wish with a snap of your fingers. At first glance this sounds unambiguous, but on second thought, what does "any person you wish" mean? How would you go about selecting a particular person that does not exist? You would have to search in a pool of infinite possible persons, all of which look exactly the same to you, and to each of which you not only lack direct access, but also indirect access, as they do not even exist. For existing persons, we also do not have direct access, but at least we have indirect access, identifying them through the effect they have on this world. We can see their bodies move, hear their voices when they speak, etc. and we infer confidently that these effects originate from a particular person which is just like our own selves. And we can track that person by following the continuity of its trail of effects on the material world which is accessible to our senses. But this is not available for non-existing persons. They are completely out of reach.

Supposedly then, you decide to exercise the power with which you have been bestowed. How would you go about doing it? You would presumably think in your mind that you want such and such a person — e.g. intelligent, emotional, curious etc. (the peripheral qualities) — but could you perceive in your mind *who* that person would be? You can imagine being that person, but this is just projecting your own self onto him/her. It is merely an illusion of direct access, not real direct access. It is essentially you, not that other person. So, suppose that after you have carefully thought and decided about how you want the new person to be, you snap your fingers, and the new person is created. Do you think you could answer the question of that person, "why is it *me* that was created"? Not really. You just imagined a person, not specifically*that* person (perhaps you imagined your own self in place of that person, with his/her peripheral qualities). You could not have determined that it be him/her that was to be created rather than some other person among an infinity of possibilities. Therefore, the real work, the selection/determination of the person, has not been done by you but by some other, unknown factor or agent. And what factor could that be? There are no possible candidates. It could not be another person or consciousness (elemental or not), for the same reasons that it cannot be you. And it cannot be an inanimate material/physical factor because all of these are equally neutral towards all possible egos.

To look at this argument from the other side, imagine going back in time to before you existed, and imagine a universe without you, to whom you are inaccessible, detached, from which you are isolated. You are in a "third-person" relationship with all elements of that universe. How could that universe produce you specifically, if it does not have any access to you? If it cannot distinguish you from any other, infinite in number, possible persons? If you are the only one who can tell the singular difference between you and anyone else, but you don't even exist back then?

Each person is unique, and its uniqueness, what distinguishes him/her from all others, from all other things in the universe,

can be seen, experienced and accessed only from within him/herself. This uniqueness is the person's identity, who he/she is[17]. It remains the same through any change in peripheral qualities, ageing, sleeping, even death. For, if I ceased to exist, my first-person perspective would still be inaccessible to anyone else; and any new person that were brought into existence thenceforth would either be someone else, with their own unique first-person perspective and identity, one that is disjoint from what mine was and that could exist alongside me if I had not perished, or it would be me, the unique and non-duplicable person that I am, brought back into existence, having the same first-person perspective as before, even if I have no memories of my previous life. My first-person identity, I myself, is a unique place in the universe, in all of reality, reserved only for me no matter whether I am alive and conscious or not.

This identity is the main ingredient for creating a person and this ingredient exists only in that person itself, and it does not exist prior to the existence of the person. So, it is impossible for an ego to be brought into existence solely by factors/agents outside of it. I could not have been brought into existence solely by the action of an agent that is outside of me and who therefore has a third-person relationship with me. It is not possible for the world outside of me to somehow know me well enough to select me, even before I existed, among an infinity of indistinguishable potential others and create *me*, when the only one that can know/access *me* directly is my own self. Any ego that is not currently part of the universe is impossible to be brought into it. And yet I now exist, whereas apparently there was a time when I did not exist. How can we get around this problem?

4.5. Solutions to the problem

So, we can summarise the creation problem thus: At the core of the essence of a person is that he/she is a centre of first-person perspective, of consciousness, an inner world which is unique to that person and inaccessible, isolated, invisible, to all the rest of the world. Between that person and the rest of the world there is the impenetrable barrier of privateness. Any attempt to "explain" how a person is brought into being should address this most important, distinctive aspect of a person: who that person is; why it is this person and not someone else that was brought into existence. But this cannot be explained with reference to any elements/factors/agents that existed prior to the creation of the person, since all of them are outside of that person, and the particular "selfness" of that person, which is what we are trying to explain, cannot be found even to the slightest degree in any of them. In fact it can be found only in that person itself and nowhere else in the universe, so the rest of the universe cannot account for bringing that person into existence. The creation of persons therefore appears to be miraculous. I will suggest two possibilities for overcoming this problem.

One solution to the problem might be that persons/egos are not created; they have always existed. This is reminiscent of conservation principles in physics, and may be compatible with beliefs about reincarnation (metempsychosis) such as those held by Plato [36] and by a number of Eastern religions. Of course, this solution does not completely satisfy an inquiring mind, as it does not consider an ultimate source of persons, a source that may transcend time if time is assumed to have no beginning. But then again any theory trying to explain reality is bound to run into the problem of recursion: if something is explained in terms of something else, then the question automatically arises of how that "something else" is itself explained. This solution also has to face the problem of the impossibility of traversing an infinite amount of time (see, e.g., [[37], §25.2] or [[38], §8.3]): if I have always existed, then I have lived through infinite events, infinite equal intervals of

time (e.g. seconds, hours or millennia), until I have arrived at the present moment. But this does not seem logically possible, or even intelligible. For example, it means that I could, if I wanted to, have recited all infinitely many natural numbers by now, even if I recited only one such number per millennium. The problem arises from the fact that I am here now, and therefore an infinite number of time units must have already passed — with the notions of "infinite" and "passed" not being particularly compatible with each other. On the other hand, one may acknowledge that this seems indeed impossible given our sense of the flow of time as being (approximately) constant, but what if, in a higher form of existence, we experience time entirely differently, in a way that is currently inconceivable? In my opinion, our ability to understand reality has limitations and we should not place our own rationality at the seat of the ultimate judge; otherwise we will run into insolvable problems no matter what we assume to be the foundation of reality, including God [39], as the present hard problem highlights.

However, my preferred solution is instead that egos are created by an omnipotent Mind, for whom nothing is impossible. From the views I have expressed so far, the reader will have figured that I believe in the supremacy of the mental over the physical; indeed, I believe that a single ego, a person, is incomparably more valuable than all of the impersonal universe. Quite the opposite of what physicalism preaches, I believe that it is the physical that depends on the mental for its existence, and not the other way round (this is a kind of monism usually referred to as ontological idealism). Nevertheless, we all know from experience that while the physical world has a great influence on our lives, our influence on the physical world is rather more limited. Furthermore, our experience of ourselves teaches us that we have limitations in our capacities, we are weak, not completely autonomous. We cannot ourselves be the source and foundation of reality, despite being infinitely more valuable than the physical world. We have to live inside a world with restrictions and laws imposed on us from outside of us. So, if the source of all reality must be mental but it is not us, this leads to a supreme Mind that is the source of all reality — called God in religious terminology (physicalism also posits an inexplicable primitive foundational principle of reality, but it is impersonal, inanimate, unconscious, instead). This Mind is like us in some respects, but much different in others. Since He[18] is the source of everything, of *all* reality, he cannot be bound by its rules as we are. Therefore, nothing is impossible for Him, even the creation of egos out of nothing[19].

In particular, concerning persons, my belief is that God has *direct access* to each of us, in a mysterious way that transcends our epistemic abilities. He is the only one, besides our own selves, that has such access. In fact, I would say that He knows us (in the full sense, not as abstract, descriptive knowledge) better than we know our own selves. It has to be this way in order to overcome the hard problem of creation, but this is not the main reason why I believe so. Another reason is the omniscience that is expected of a God, the source of everything; this omniscience is often interpreted as "God knows all true propositions" [40], [[37], Chapter 28.1] but this sounds like very theoretical, superficial knowledge, like someone knows something when he reads about it in a book. This kind of knowledge is knowledge in name only, and is certainly not worthy of God. If I know that the proposition "John is in pain" is true, but I've never experienced pain myself, what kind of knowledge is that? Even if I have experienced pain myself, I cannot claim to have perfect knowledge of John's being in pain, in the sense required of God, unless I am experiencing it from John's own first-person perspective. God must know things directly and fully, completely, from the inside. This includes our inner states, even if sinful, and it includes our particular first-person perspective. A creator who knows his creations only theoretically, descriptively, propositionally, or even by imagination (imagining himself in another's place) is not a creator in the full sense of the word

but merely an assembler of things from pre-existing elements of reality, a reality that exists independently of him with its own rules, laws, essences, properties etc. which the creator exploits to achieve his creation. But if all of reality originates from the Creator, including these rules, laws, essences, properties etc.then He must know them completely and directly, from the inside[20].

Another reason, the main one for me, is God's goodness combined with his omnipotence. If persons are the most valuable thing in the universe by far, and we, the limited persons, love them and want to empathise with them, would not God want to empathise with us completely? Which is to know us completely, from the inside, directly; to form some sort of union. In fact, I would say that union with him is what he had in mind when he created us, the ultimate goal. The intrinsic value of persons translates into love for them, and love desires union. Perhaps in the afterlife the impermeable barrier of privateness will be somewhat relaxed even for us, through the intermediation of God who has direct access to each of us. On the other hand, it seems to me that it is also love that dictates some privateness as we do not want to intrude into other person's lives and inner world.

Finally, I think that direct knowledge of each person is also a prerequisite for God to be able to pass fair judgement on them.

5. Free will

5.1. Introduction

So far, in Sections 2 and 3, we saw that physics cannot account for all of reality but that mentality exists fundamentally. Then, in Section 4, we saw that mentality is not a property of the material world, but it is exhibited by minds, persons, who are simple, fundamental substances (in fact, in my opinion they are the prototypical substances, other than God — but in fact they are prototypical precisely because they are God-like). But so far the arguments put forth do not rule out the possibility of epiphenomenalism, that the physical world, including our own bodies, goes its own way according to its closed set of laws, while the mind simply follows along, in a weak-supervenience relationship with the brain, under the illusion that it has its own free will and can make thoughtful decisions that impact the world based on reasons and choice, whereas all along what determines our behaviour is the chemical procedures in our brain according to the laws of physics and chemistry[21].

Epiphenomenalism is a requisite for any theory that wants to maintain that the physical realm is causally closed, including physicalism and, as far as I know, panpsychism. This means that knowledge of the initial state of a human body, and of the time history of physical stimuli that it receives from its environment, allows us to predict the future state of that body down to the last detail by applying the physical laws alone. There is no mental causation (e.g. Cartesian interactionism). Theoretically then, if epiphenomenalism is true, someone's actual behaviour can be predicted precisely by an atomistic numerical simulation of his/her whole body like that described in Section 2, where the computer program is fed with the initial state of the body (the particles it consists of and their initial locations and velocities), the boundary conditions (the physical forces acting upon the body from its environment), and the laws of physics such as Newton's laws of motion and Maxwell's laws of electromagnetism. From a psychological, conscious, "intentional stance" [24] point of view that person

has reasons to behave as he/she does — e.g. has to pick up the kids from school, or studies a book to learn about quantum mechanics out of curiosity or to get a degree in physics, or buys a present for a friend that she loves to make her happy. But these reasons play no role whatsoever in our simulation and are completely irrelevant; closure of the physical laws means that the behaviour of a physical system such as our body is governed entirely by the physical laws, of physical/mathematical character, which have nothing to do with "reasons", such as those mentioned. A slight complication arises in case that quantum mechanical effects, and their associated randomness, are not negligible in governing the macroscopic behaviour of the body. In this case, our simulation would not predict a definite behaviour, but a range of behaviours, each assigned a different definite probability. Some think that such a randomness leaves open the possibility that physical causal closure is compatible with free will [41][42]. However, under this “quantum mechanical randomness” scenario, our will, despite not being deterministic, would also not be free; it would more appropriately be called "random will". Reasons would still play no direct role in shaping our behaviour, and our choices would not be freely decided by ourselves, as persons-substances, but would be the result of a combination of insentient, physical, determinism and randomness. To predict our behaviour, we would still solve equations that take absolutely no account of any reasons we may have for acting as we choose to act, but would involve only physical quantities such as positions, forces, charges and velocities, and probability distributions.

Our intuition is that we have what in philosophical terminology is called "libertarian" free will, that is, that we make our own choices which are not predetermined either by the laws of physics or by anything else. It is us, as minds-substances that choose freely. Free will is, in my opinion, not something that can be defined precisely in terms of other concepts; it is a fundamental property of ours, as mental substances, that we know directly from first-person experience. And, of course, what we experience in ourselves we project also to others, and hence this intuitive perception of freedom lies at the foundations of our regarding people as morally responsible, judging them as worthy of blame and praise, but also at the foundation of all our interpersonal relationships. When one holds towards us or towards others a friendly or hostile stance, cares and respects them or is indifferent and self-centred, we think that it is their own choice, that they do so freely, and we form our own opinion of them based on this intuition. In relationships between friends, between parents and children, between spouses, etc. one intuits that the other party freely chooses to love them and that they could not love them if they so chose, not that they are merely programmed to love them, as epiphenomenalism claims. I think that an eloquent description of free will is the following by Chisholm:

> *"[W]e have a prerogative which some would attribute only to God: each of us, when we act, is a prime mover unmoved. In doing what we do, we cause certain events to happen, and nothing – or no one [not even God] – causes us to cause those events to happen". ([43], text in brackets is mine).*

I would like here to supplement the above description of free will with the observation that we can only directly cause events within our own minds, whereas events that occur in our bodies as a result of our mental decisions are only indirectly caused by us, through the mysterious mechanism that binds mind and brain. One can have free will even if their body is paralysed and they cannot engage in any physical action. Making a decision to act physically is itself a mental event, often the last one in a chain that consists of rational thoughts, which we can direct at each stage by making

choices, assenting to something, rejecting something else, questioning or inquiring into something, etc. So a decision to act on some matter is in fact the result of a thought process that can involve multiple micro-choices, micro-decisions along the way. Our exercise of free will is not something that we perform only occasionally, but is a fundamental, ubiquitous aspect of mentality. Free will cannot be separated from the first-person perspective, it is an integral part of it. Over a period of time, or over our entire lifespans, we can volitionally settle on certain views and beliefs about reality, through such volitionally initiated and directed thought processes, thus shaping our own characters. Of course, one can change his/her mind, even from views he/she has previously settled upon, and change his/her character. A decision we make may or may not involve bodily action (e.g. the decision whether to believe something or not does not have a direct and immediate effect on our bodily behaviour). All of these mental motions, being voluntary, can carry merit. Thus, if one assents to adopt the belief that people in need should be aided, this decision carries some merit, even if he/she has not actually helped anyone yet; and if one assents freely to accept the Nazi ideology, this deserves some blame, even though he/she may not yet (or ever) put the Nazi tenets to practice. Similarly, in our interpersonal relationships we feel gratitude for the volitional love that someone feels for us, even if they did not practically externalise this love through physical action, such as aid in time of need or gifts etc. In fact, even if such action is exhibited, it is the mental motions from which it derives, i.e. again that person's volitional love for us, that moves our own volition to love them back, with the actions serving to reinforce our perception of the fact that they love us and providing a measure for that love. For, if we perceived that these actions were not performed out of love but for reasons, say, of personal interest (e.g. expecting something in return), we would not be moved.

So, summarising, free will is very hard, maybe impossible to define precisely. It seems fundamental and not analysable into other notions, but we are most intimately familiar with it by introspective experience since it is always active, to various degrees of intensity, synchronously with consciousness and cannot be separated from it. The traditional definitions of free will as the ability to do otherwise or as the property of being the ultimate originator of one's actions are insufficient, since a fundamental particle could have behaved differently due to quantum mechanical indeterminacy, while an electron is the ultimate originator of its action to repulse other electrons (setting aside God), but these properties have nothing to do with free will. Free will without consciousness is unintelligible, but I would also argue that the converse is also true, that free will is a defining part of consciousness. In my opinion, free will is a fundamental aspect of personhood, one of the defining elements of the first-person perspective.

5.2. Reasons and causes

A person has reasons for willing, deciding and acting as he/she does. The fact that free will is indeterministic does not mean that it is random or irrational. Reasons have to do with rationality, but also with morality, which does not ultimately derive from rationality although it is always manifested in combination with it. Someone can be good or bad, fair or unjust, noble or wicked, irrespective of how intelligent they are. In my opinion, morality is not fundamental but derives from the innate value of persons; Jesus expressed this eloquently when he said that all the rules of morality ultimately come down to loving God and one's fellow person (e.g. Matthew 22:36-40; see also Matthew 7:12). Persons are existences, small gods, and their infinite value derives from their likeness to God, i.e. the first-person perspective — being someone rather than something. Therefore, while the physical laws are creations of God, morality and its "laws" are not creations but

derive from the very nature of God, which is similar to our own. But it is not a matter of intelligence whether we will choose to put our own selves first and pursue self-interest or if we will (at least try to) love others the same as ourselves, recognising the infinite value of all persons and empathising with them, putting our selves in their place. Whether we will do one or the other is a matter of free choice, and in my opinion, even though we exercise our free will throughout our daily activities, the primary application and role of free will is in in relation to making moral choices. I would go as far as to say that coming to choose freely good over evil, to put it simplistically, is the meaning and purpose of this life.

As noted in Section 3, persons are both conscious observers of reality and actors who contribute to its shaping. Our actions are based on our understanding of reality, which allows us to predict the consequences that our potential action or inaction will have and to evaluate the alternative choices. Some confuse rationality with determinism; they think that our decisions are determined by the reasons we have for making them. In this view, our reasons for acting are the causes of our actions [44], not us — we are merely the vehicles through which these reasons cause the actions. Furthermore, they think that indeterminism is synonymous with randomness, so that if determinism were false then our decisions, not being determined by reasons, would be completely arbitrary and purposeless. In my opinion, this is a complete distortion of reality. When someone faces a dilemma, this means that there are valid reasons for taking each of the possible alternative paths of action, and it is up to the agent to choose among them. The reasons do not determine which path is to be taken. This is especially apparent when the dilemma is a moral one, because then there is no objective, rational weight that can be assigned to the various reasons for making the moral or immoral choice. If the dilemma does not concern morality, then rationality, intelligence, logic may indicate that some reasons should carry more weight than others, in order to achieve a certain goal (e.g. when one thinks about whether to go somewhere on foot or by car, about which brand and model of mobile phone to buy, about where to invest their money, or about what actions to take to advance his/her career); nevertheless, even in this case the reasons by themselves do not have any apparent determining force, only a motivational one.

For example, suppose Giannis, on his way to work, finds a wallet full of money. He can either keep it or hand it over to the police. As he is contemplating what to do he may think that the owner worked hard for that money, that he may have a family in need of it, that returning the wallet will strengthen the owner's faith in humanity and promote goodness, etc. On the other hand, he could instead think that the owner should have been more careful and it is his fault that he lost the wallet, that he may be rich and not need the money as much as Giannis does, that right and wrong are mere human conventions and do not exist objectively in the natural world, that each person must prioritise their own interests or act on survival instinct as is done in the animal kingdom, that life is unfair anyway and people working hard may earn less than others who work little or not at all (e.g. who have received a large inheritance or won the lottery), etc. Which thoughts he will concentrate on is up to him, it is his choice. Whatever he decides, the determinists will say that it was the reasons behind the choice that determined it. However, there are reasons for both choices; whichever choice he makes there will be reasons for it, the reasons that he freely chose to value the most. The same reasons will be available also for a person that makes the opposite decision than Giannis, showing preference for another reason. Therefore the reasons do not determine the choice, it is Giannis himself who determines it. Furthermore, the fact that Giannis may have contemplated about these things since many years prior (e.g. since childhood) so that he has settled, say, to try to always do the right

thing, is not proof that he behaves deterministically since it was he who shaped his own character on his own free will, as argued previously. By exercising our free will in small and big matters throughout our lives we steer the development of our character, and hence the decisions we make later in life are consistent with the worldview we have built through the years. Hence, it could happen that Giannis finds a new lost wallet every day and always non-deterministically turns it over to the police. Indeterminism is not synonymous with randomness.

In fact, it seems implausible that reasons could ever be actual determining factors of behaviour, whether the world is deterministic or not. A "reason", in the current context, is founded on an understanding of reality, on the meaning of reality, on an understanding of how it is, how it could be, and how it works, its "mechanics"; this understanding is purely mental and not physically explainable, as was argued in Section 3. For that understanding to become a "reason" it must be combined with the agent's free will: the agent on his/her own decides to be the originator of an action, to make a change in the world, informed by this particular chosen understanding. The understanding alone does not have any causal power by itself. In libertarianism, it can merely provide motivation for someone to cause something on their own. In epiphenomenalism (e.g. in physicalism), understanding, deciding and acting have physical substrates, i.e. biological structures and events in our bodies (brains), which monopolise the causal power leaving nothing to the mental understanding. Let us call these physical substrates PU (physical substrate of understanding), PD (physical substrate of deciding) and PA (physical substrate of initialising bodily action), and their mental counterparts MU (understanding), MD (making a decision) and MA (the conscious commanding of our body to act). Then, according to epiphenomenalism, PU causes both MU and PD; PD in turn causes both MD and PA; and finally, PA causes both MA and the contraction of our muscles so as to perform the decided action. The causal chain PU → PD → PA is precisely predictable by molecular dynamics simulations (i.e. by the laws of physics) while the causal relations PU → MU, PD → MD and PA → MA are not derivable from the physical laws, as discussed in Sections 2 and 3. PU, PD and PA caused everything, and our intuitive perception that it was the mental processes (MU, MD and MA) that caused the action, i.e. that it was our understanding of reality that motivated us to decide to act in such a way for the reasons we chose, is illusory, as physical causal closure requires. Everything happened according to physical meaningless and purposeless laws. In principle it would be possible to simplify the biological substrates by removing those parts of theirs that give rise to mental phenomena, leaving the rest intact (so that PU causes only PD but not MU, PD causes only PA but not MD, and PA sends signals to the muscles but does not cause MA), so that the body would perform the exact same action as before when excited by the exact same external stimuli but absent a mind, leaving what resembles a philosophical zombie, a biological robot, or perhaps an animal if Descartes is right. Hence, removal of parts of Giannis' brain could result in a mindless zombie-Giannis who nevertheless picks up the wallet and hands it over to the police, as the behaviour-related neural circuitry that starts from the eye receptors that receive the light reflected from the wallet and ends in the muscles that pick it up and walk over to the police is left intact.

So, anyone claiming that reasons are causes, not in a loose sense of providing motivation to an agent but in a strong sense of determining him to act in a certain way, must provide the missing link of where the determination comes from. If our behaviour is indeed determined, then there must be other causes that cause it, because reasons as defined here do not explain how the determination arises. Therefore, if the explanation offered for why Giannis decided to return the wallet

is simply that he wanted what is best for its owner, without that being further analysable, then this explanation is reason-based and makes sense only if libertarian free will is implied. If, on the other hand, one claims that Giannis' behaviour was determined by his brain's chemistry, then it is misleading and incorrect to claim that it was the reason (Giannis' preference to do what is best for the owner) that determined his behaviour, because it really had nothing to do with it — it was all chemical mechanics.

Suppose that someone, like Giannis, is faced with a dilemma of choosing between A and B, and that (physical causal closure being true) this choosing amounts to a particular neuron in that person's brain firing or not. But whether or not it fires depends on whether enough neurotransmitter ions reach its receptors, which can be determined by solving a few equations that involve things such as the locations, masses, charge distributions, conformations and orientations of the ion molecules, of the receptors, and of surrounding molecules in a small neighbourhood. One can deduce whether that person will choose A or B by solving the equations of electromagnetism and Newton's second law, while what choices A and B mean and what reasons that person may have to make each choice are completely redundant information. An objection may be that while the situation is indeed so, i.e. the direct and immediate cause of the choice is physical, the reason why the neurons, neurotransmitters etc., are so arranged in the first place is due to the reasons behind the choice; that is, the mechanics of the behaviour is reasons-based, and physical causation is but the vehicle for implementation. But let us go backwards in time by a fraction of a second δt. Is the current physical state $S(t)$ of the neuron, neurotransmitter and surrounding molecules not explainable by reference to the past state $S(t - dt)$ and to the laws of physics alone without any reference to reasons, meanings, purposes, and any meaningful element of human mental life whatsoever? And that state in turn, is it not in exactly the same manner fully explainable in terms of the state $S(t - 2\delta t)$ at a moment earlier and the physical laws? And so on to the beginning of the universe. All along then, the history of the universe has not been shaped in the slightest by anything other than its initial state and the laws, with reasons, meanings, purposes, intentions, desires, hopes, fears, will, decision etc. not playing the slightest role.

The epiphenomenalist may also pose the following objection. He/she may grant that reasons defined as mental understandings associated with decision-making do not by themselves have determining power, but argue that the actual facts that the understanding refers to do have such power. Therefore, when we attribute someone's behaviour to a certain reason, what we mean is not that the mental understanding of some facts about reality causes this behaviour, but that these facts themselves cause both the understanding and the behaviour (in typical epiphenomenalistic fashion). For example, when Giannis finds the wallet it is a fact that the wallet is in front of him and the light reflected on it reaches his eyes which in turn causes a series of chemical processes in his brain. However, there are serious problems with this view, namely that the "facts" on which our reasons for behaving are based are often (or rather, usually) not physical facts. For example, they may be absent (Giannis has never seen the owner of the wallet), hypothetical ("the owner of the wallet may have a family in need of the money"), psychological ("the owner's faith in humanity will be strengthened"), or refer to ethics or other non-physical aspects of reality ("it is the right thing to do", "it is for the owner's benefit"; ethics will be discussed further in Section 6). Such "facts" cannot interact with Giannis' brain in any way and therefore cannot cause anything to happen there.

Sure, the physical realm and its laws are complex enough such that they can produce very intricate behaviour, which can perhaps mimic exactly the behaviour of a libertarian agent who acts based on reasons. Even so, while the observed behaviour would be the same on the outside for the two cases (physicalist and libertarian), their foundations would be incommensurable with each other. For example, consider a physicalistic analogue of caring for another, as Giannis cares for the owner of the wallet. Let us assume for the sake of argument, similarly to Descartes, that animals lack minds, that they are automata, biological machines, or "philosophical zombies" in the terminology of Section 2. Suppose then that a wild dog is attacked by another animal, and the rest of the pack comes to its aid. Why would this happen? The meaning of "why" is, in this case, different than the "why" that asks for a reason. Whereas in the case of a libertarian agent "why" has to do with reasoning, with the meaning of reality, the appropriate version of “why” in the present case has to do with physical causes. So, actually, the behaviour of the dogs is completely determined by the biochemical processes that occur in them; it is determined by attractive and repulsive electromagnetic forces between molecules and ions in their bodies, and most importantly in their brains, governed by the laws of physics. After evolving for many generations, the physical structure of the dogs is such that, when the patterns of light and sound waves that are generated when one of their kind is attacked by another animal reach their sensory organs, the corresponding generated patterns of neural signals will travel from there to their brain, where they will trigger a deterministic chain of chemical processes that will ultimately send signals down to the their muscles and cause them to move in defence of the dog being attacked. It is a predictable, purely physical mechanistic process that results entirely from the physical structure of the dog and the physical laws, and has nothing to do with reasons, thought, and an understanding of reality. We could ourselves design and build electro-mechanical robot "dogs" that behave in the same way. The biological dogs have this physical configuration because this behaviour increases their probability of survival, and so their ancestors, in which the random mutation responsible for this behaviour first occurred, survived to reproduction and passed their genes on to subsequent generations. On the contrary, other species such as, say, gazelles, do not exhibit this behaviour because those of their ancestors who developed such a mutation perished, as this behaviour proved detrimental to their survival since they are not well equipped for fighting another animal. So, modern gazelles' brain structure is such that the corresponding chain of chemical processes will send signals to the muscles to flee rather than fight.

Due to privateness, a mental observer of the dogs may wrongly apply the "intentional stance" of Dennett literally and assume that the dogs have minds, and interpret their behaviour as originating from the reason that they care for the dog being attacked and they want to save it. However, the true origin of their behaviour has nothing to do with this; it is completely mechanistic, physical, mindless, meaningless, and purposeless (it works by increasing their chances of survival, but survival is, of course, not a "purpose" in a physicalistic setting, unless we speak figuratively). Hence, if our behaviour is merely physically caused like that of the dogs then the reasons we think we have for doing anything are illusory.

This would mean that the world is more simplistic, mechanical, purposeless and meaningless than we think, and that our view of the world as reflected in our reasons and the understanding on which they are founded is subjective and fake, existing only in our minds, as a side effect of the physical processes that occur in our brains. This includes our philosophical theories (including physicalism) which, being a product of our reason, would seem most likely to be

completely wrong, based on illusions. Hence the criticism against physicalism that it is self-defeating known as the "argument from reason" [28][45]. Although this argument is directed against physicalism, it is clear that any epiphenomenalist theory, including panpsychism, faces the same problems.

The only way for the epiphenomenalist to escape this conundrum is to solve the hard problem of Section 3 which would allow that the physical processes in the brain map naturally and exactly onto the mental thoughts, the reasons, which would require that the seemingly meaningless and purposeless physical structures in the brain have inherent meanings. In other words, the mental meanings and the corresponding brain structures and processes must be two sides of the same coin, they must be identical. In this way, it would actually be the reasons that cause the behaviour, through their material form in the brain structures, which have physical properties such as mass and charge, can exert force etc. It should be noted that we are not talking about a mere contingent correlation between meanings and brain structures or processes; such correlation is already assumed in the epiphenomenalist model that faces the aforementioned problems. The correlation must not be contingent, but logically necessary, i.e. it should be logically apparent why a certain meaning and its corresponding neural substrate are identical. If they are one and the same, then the laws that govern one are the same as those that govern the other, and therefore the physical laws that govern the physical substrate are identical to the logical laws that govern reason; one set of laws would be reducible to the other, and the reasons we have for behaving as we do would be identical to the physical causes that cause our behaviour. Of course, this is completely implausible because the correlation between meanings and biological structures is obviously contingent, as discussed at length in Section 3. For example, metaphysics would need to be derivable from physics. Furthermore, it is logically impossible for the physical substrates to be entirely equivalent to the reasons because, as we saw, reasons by their nature do not determine, whereas their purported physical substrates do have a deterministic nature. Reasons and causes have completely different, irreconcilable characters.

5.3. Agent causation and physical causal closure

The account of free will presented above is called "agent-causal"[[46], p. 243-259], because it attributes causational powers to the agent, regarding him/her as a simple substance (the agent is a "prime mover unmoved", in the archaic terminology of Chisholm). Some attempts have been made to present this view as inconsistent, which are discussed by Pereboom [[47], Chapter 3], who, to his credit, despite believing that we do not in fact have free will, argues convincingly that all of these attempts are unsuccessful. Nevertheless, he finds the theory implausible on empirical/scientific grounds. As far as I know, there are no such empirical grounds that we know of as yet; Pereboom, like many others in the current intellectual realm, confuses science with physicalism (I use Pereboom's views as a characteristic example, but I think that they express the subconscious unfounded physicalistic beliefs of much of modern society). His objection is related to a very important consequence of agent-causality (of genuine free will, essentially): If agent-causality is true, then the laws of physics are routinely violated somewhere in our brains, or, to be more precise, physical causal closure does not hold somewhere in our brains, at the mind-brain interface. Indeed, the current state of my body (initial conditions), the external stimuli it receives (boundary conditions), and the physical laws, determine (let us at first ignore quantum mechanical indeterminism for simplicity) the future states of my body, which can, in theory, be predicted by a numerical simulation program, as

mentioned. If, then, according to physics it is determined that my arm remains lowered during the next ten seconds, but, having free will, I decide to raise it and in fact do raise it within that time frame, then this means that something took place in my brain that was not predictable from the laws of physics alone. I do not even have to engage in physical action in order for there to be deviations from physical determinism in my brain: if my thoughts are perfectly correlated with neural processes in my brain, then merely thinking freely will cause such deviations. Quantum mechanical indeterminacy does not provide a means of reconciliation between free will and physical causal closure. Such indeterminacy, if relevant at the scale of neural processes, would translate into specific probabilities of how likely I am to perform each action within a range of possible actions; if I am truly free, then these probabilities would be violated (e.g. I could consistently decide to do the least possible action, or an action with zero probability). In relation to these implications of agent-causality, Pereboom writes, for example:

> *"The most significant empirical objections to agent-causal libertarianism challenge its capacity to accommodate our best natural scientific theories. Different aspects of this type of libertarianism give rise to two such objections. First, given our scientific understanding of the world, how could there exist anything as fabulous as an agent-causal power? It would appear that our natural scientific theories could not yield an account of a power of this sort. Second, given our scientific understanding, how could there be agent-caused decisions that are freely willed in the sense required for moral responsibility? Such decisions, it would seem, would not be constrained by the laws of nature, and therefore could not exist in the natural world." ([[48], p. 69]).*

I fail to see any conflict between our best scientific theories and agent-causality (or Cartesian dualism), or rather, I find it impossible that there can even be such conflict. Scientific theories describe the entities that comprise the physical world, and the interactions between them in the form of laws *under the assumption that there are only physical entities involved* In order for two theories to conflict, they must provide different interpretations of the same phenomena, but agent-causality is not about the purely physical phenomena that scientific theories interpret, nor does it deny the validity of any of the physical laws. Claiming that it somehow conflicts with scientific theories is similar to claiming that Newton's law of gravity conflicts with Maxwell's laws of electromagnetism; they can't be conflicting, since they refer to different phenomena. Our basic laws of physics were formulated based on experimental observations that were made on the behaviour of inanimate, insentient physical systems (or, if they concerned biological organisms, they concerned aspects of them that have no direct relation with free will). Cartesian dualism and agent-causal libertarianism accept these scientific findings but also accept the existence of non-physical entities, minds or persons, who therefore cannot be studied by physical sciences and cannot be described by scientific theories. So, agent-causality can not conflict with any scientific theory. To be fair, it is possible that agent-causality will be discredited by scientific *experience*, if there ever is obtained a large body of scientific data that shows that even the finest molecular motions in people's brains are precisely predictable by the physical laws and hence the manifestation of agent-causality is nowhere to be seen[22]. However, this is not what Pereboom argues here. He rejects agent-causality because it is so "fabulous" that it cannot be accounted for by scientific theories, and because nothing that is not governed by the physical laws can exist. But the tenets that everything has a physical explanation and of physical causal closure are not part of any scientific theory but they are precisely the tenets of

the philosophical theory of physicalism, which happens to face many hard problems outside of the topic of free will, as we saw in the previous sections. It is obvious that Pereboom confuses science with physicalism. Later, he does refer to the issue of scientific data, but in a totally biased manner:

> *"One major difficulty for this strategy ... is that we have no evidence that such divergences [from physical determinism] occur [in the brain]. This problem, all by itself, provides a strong reason to reject this approach." ([[48], p. 85-86]).*

Again, I find this statement quite baffling, as, whereas it is indeed true that we have no evidence that such agent-caused divergences occur, we also do not have any evidence that they do not occur. In order to verify or disprove agent-causality, we would have to make extremely detailed experimental investigations of living brains at the sub-neuron level, and check whether all chemical motions seen there are in accordance with the physical laws or not. Possible quantum mechanics effects would complicate things even further, making us deal with probabilities. Obviously, such tests are far beyond our current technological capabilities. Yet, in the absence of both positive and negative empirical evidence on agent-causality, and despite the fact our natural intuition is that we are causal agents, Pereboom takes it for granted that it is natural to assume that agent-causality is false. This is obviously a case of prejudice in favour of physicalism, which is very widespread as discussed in Section 1.

Having said all that, it is obvious that the physical world in which we live and its laws have a tremendous impact on our lives. Although in the present Section I have emphasised the freedom that I believe we have as agents, I have done that as a reaction to the prevailing physicalist view that denies it, and I do not claim that this freedom is unlimited and absolute. The world we live in, with both its physical and non-physical aspects and facts, sets limits within which our freedom extends. Giannis would not have faced a moral dilemma and would not have made a relevant choice had not someone lost their wallet somewhere on Giannis' path on that day. Wallets and money and roads and jobs and the owner of the wallet exist due to factors that are entirely out of Giannis' control. Even Giannis' own existence did not arise from his own free will, but was decided for him. His form of existence, which obviously affects greatly his ability to exercise his free will, is also beyond his control. The physical laws and the way the physical world and our own bodies are composed have a tremendous impact on how we perceive the world and what impact we can have on it and on other people, hence they weigh in hugely in our decision-making. The environment we grew up in, the experiences of our past, what we have been taught, the views of our society, also contribute to shaping our view and interpretation of the world, and hence influence how we perceive the different choices that we have, thus contributing to shaping our reasons for acting. Furthermore, apart from our bodies which are part of the physical domain, our mental, non-physical traits are also beyond our control; for example, our sensory qualia: the way we see, hear, smell etc. The mysterious correlation between body and mind is also something imposed on us, and it acts both ways so that by exercising our free will we can cause events to happen to our bodies, but events in our bodies also have an influence on our minds. Hence it is much more difficult to decide to quit smoking (and going through with one's decision) than to decide to regularly eat pastry. And of course, degenerative brain diseases, and ultimately death, obviously have an impact on our decision-making. These are all quite strict limitations, but nevertheless I do believe that these limits leave room for us to make our own choices. Furthermore, it seems that tasks

that we repeat are "memorised" by the brain in corresponding neural circuits so that after learning them they are performed in "autopilot", where only a bare minimum of input is required of free will and most of the work is performed physically by the brain. For example, walking, running and climbing stairs seem trivial, but the motions executed are in fact complex and many muscles are involved; nevertheless, we need not preoccupy our free will with these individual motions because they are automatically taken care of. And the same occurs with more complex activities, such as driving — in fact it seems to me that perhaps all our activities include some degree of physical "autopilot". From all these considerations, I would expect most, but not all, of what occurs in the brain to occur in accordance to physics.

5.4. Some arguments against epiphenomenalism

So, one reason to doubt epiphenomenalism is our introspective intuition that we have free will. Another may be the suspicion raised in Section 3 that the mapping between the mind and the brain is incomplete (weak supervenience does not hold). If this is so, then the precise contents of our thoughts would not be fully reflected onto physical processes in the brain, and so whatever is caused by these thoughts (e.g. other thoughts, and possibly eventually physical action) would not be caused physically (since the physical substrate is missing) but mentally. In Section 4 it was indeed shown that weak supervenience does not hold; of course, it was shown to fail in the mapping of persons to bodies, which is not directly related to epiphenomenalism, but nevertheless it raises doubts. Another argument against weak supervenience of thoughts on neural processes is the aforementioned "argument from reason" [28][45]. Our thoughts express reason, rationality, and understanding of reality; we assume that our mental access to, and understanding of, reality is direct and perfect. This does not mean that we do not make mistakes, but it means that if we think things over better, or if someone explains the situation to us, we will see beyond our mistake into the truth in a way that makes sense. In other words, there is an objective, absolute standard of rationality that characterises the world and we have the capability of reaching it. The theories we conceive about reality, including physicalism, are formed under this implicit assumption — that we are rational minds capable of getting to the truth. On the other hand, our brains function physically, according to laws that have to do with attractive and repulsive forces between ions, molecules etc. and which by themselves have no relation or affinity with the contents of our thoughts. Even if it were possible that a physical language of thought existed (if we overlook the insurmountable difficulties of Section 3), it is extremely unlikely that a physical system with processes occurring according to physical laws could match the truth of reality perfectly. It is impossible for there to be an inherent direct physical mechanism to force a perfect match between the physical structures and processes in our brains and the objective, rational character of reality. Our brains were formed by the indirect long evolutionary trial and error procedure whose "goal" is survival, not understanding — understanding is a byproduct, a means at best, to which evolution is not committed; it occurs by accident and there is no guarantee that it is perfect. It is suggested that a long evolutionary process driven by natural selection could match the "workings" of the brain with the "workings" of reality, because understanding reality would be beneficial to our survival. However, on one hand it seems unlikely that evolution will have gotten it right all the time, i.e. it seems more likely that only a partially correct understanding of reality would have emerged. And on the other hand, even when evolution got it right, this would happen contingently, our understanding would not be real but merely a fortuitous coincidence between neural workings and some objective truth of the world out

there to which we would still not have direct access. Furthermore, if one accepts the objectivity and reality of things such as good and evil, right and wrong, or maybe even beauty, ugliness, cuteness, funniness, seriousness etc., then he/she is faced with the problem that these can have absolutely no effect on the physical brain: they can not pull or push ions, exert forces, they don't have mass or charge, etc. Evolution is a physical trial-and-error procedure, and to produce adaptation of an organism to elements of its environment it must be the case that the organism physically interacts with these elements. In particular, if evolution is to produce understanding of elements of reality, the brain must be able to physically interact with these elements, but when it comes to concepts such as good and evil it cannot because they are not physical. So, a physicalist is forced to think of them as mere human conventions, constructs, in irrealist terms. To conclude, it seems that the physical structure and operation of the brain is subject to constraints that do not seem to apply to the mind and its mental capacities. If this is true, then supervenience between the two is not possible.

5.5. The paradox of predictability

According to Spinoza, an early hard determinist,

> *"[H]uman beings are mistaken in thinking they are free. This belief consists simply of their being conscious of their actions but ignorant of the causes by which they are determined. Their idea of their freedom therefore is not knowing any cause for their actions[23]". (Spinoza 1677, Ethics 2P35S [[49] p. 73]).*

Science has made significant progress since the time of Spinoza, and the features of the physical world that modern like-minded philosophers believe to be the determining causes of our behaviour are now known and understood to a large degree: they are the fundamental laws of physics, which govern the behaviour of the countless fundamental particles that compose our bodies. But then the following questions arise naturally: if we know the causes that determine our actions, and we know the mechanics of these causes, can we predict our own actions? (Even if quantum-mechanical indeterminism does not allow us to make a precise prediction, it still determines the probabilities of our potential actions, and these probabilities are predictable). After all, nowadays numerical simulations (predictions) of the behaviour of physical systems are routinely performed in many, if not most, scientific and engineering disciplines, exploiting the deterministic nature of the physical world and its law-governed behaviour. And if indeed we can predict our own actions, which seems reasonable if we are just physical systems, what will happen if we choose to act contrary to those predictions, i.e. to act differently than what Spinoza's causes dictate? Our power to do just that seems equally reasonable, and having this power would disprove epiphenomenalism and prove that we have free will[24].

Unfortunately, while based on an interesting intuitive idea, this argument fails to deal a conclusive blow to epiphenomenalism. The idea is based on two pillars. First, on the fact that we know from introspection, intuition and experience, that we have the intimate ability to choose between available options; sure, oftentimes we can feel strongly pressured or pulled towards a certain option (such as the option to smoke a cigarette, for a smoker), but this does not completely take away our freedom of choice. And secondly, on the fact that the physical world operates in a relatively predictable, mechanistic way, that is knowable to us. Combining these two ideas, it follows that if we are physical systems

then the mechanics of our behaviour is predictable, and our behaviour is knowable in advance, but if we know how we are supposed to behave we can choose to behave differently, which would disprove the assumption that we are physical systems. So, where does this line of thought go astray?

The weak link is the assumption that one can know what the physics of his/her body dictates for his/her future behaviour, which is a prerequisite for disproving that his/her behaviour is all physics. This assumption, despite seeming quite plausible (at least in theory) at first glance, runs into serious trouble upon further consideration. Firstly, the structure of our body is too complex for the physical processes occurring therein to be tractable by an unaided human mind. This may, at first sight, seem to be a contingent fact that does not affect the theoretical substance of the argument. But a physicalist may retort that this fact is actually not contingent at all, and is easily explained if we accept that the "mind" is actually physical processes in the brain, and the inability of the mind to keep track of the complex processes in the body is nothing more than the inability of the brain to represent the future state of its own self, which is due to the impossibility of self-prediction for any physical system[25]. Furthermore, it turns out that it is, in general, not possible to use an external aid, such as a computer equipped with simulation software, to acquire the knowledge of one's own future behaviour according to physics. In order to acquire such knowledge, one must physically interact with the computer, which effectively invalidates the assumptions on which the prediction was based.

These issues, which are referred to as the "paradox of predictability"[50], are not particular to the prediction of the behaviour of minds but under certain conditions pertain also to the prediction of the behaviour of perfectly deterministic physical systems. An extensive discussion of this "paradox" is planned for the future in a separate publication. These issues dismiss the possibility that the argument under discussion poses a hard problem for epiphenomenalism. Nevertheless, I personally do not completely dismiss the argument as devoid of any value, but I think that further contemplation can be beneficial for gaining deeper understanding of the issue of free will. In particular, it does not seem to me that we have a clear case of self-prediction, due to the duality of the characters of mind and brain. The case for epiphenomenalism will collapse if it turns out to be possible for one to become aware of what he is supposed to will according to physics.

6. Ethics

Finally, I would like to discuss some implications of physicalism, and of any theory that assumes physical causal closure (epiphenomenalism), on the ethical aspect of human life. Strictly speaking, ethics does not pose hard problems for physicalism, or for any other theory about the nature of the mind. Many of these theories indeed have implications that are incompatible with our intuitive view about the ethical side of the world; for example they may imply that good and evil, or right and wrong, do not exist. Nevertheless, no matter how disruptive these implications are to personal and social life from both theoretical and practical perspectives, they do not constitute inconsistencies for the theories of the mind themselves. If these theories are true, then this may mean that reality is fundamentally horrid, meaningless, etc. but we would just have to live with that. Nevertheless, these implications may motivate one to take views such as Cartesian dualism and free will libertarianism more seriously, in fact to hope that they are true rather than alternative theories. So, in

this section I will present briefly some such implications of physicalism and other epiphenomenalist theories.

6.1. Ethics and free will

Ethics has a close connection with free will. Hence, let us begin by discussing some ethical consequences of epiphenomenalism, of the view that our thoughts, decisions and behaviour are governed entirely by the physics of our body, that is, by physical determinism possibly complemented by some degree of physical (quantum-mechanical) randomness. Most philosophers nowadays subscribe to this view; they are either hard incompatibilists or compatibilists. The difference between the two is that the former (rightly, in my opinion) recognise that if our behaviour is ultimately determined by physical causes alone then this absolves us of any real guilt for wrongdoing, and deprives us of any right for credit for any noble actions we "choose" to perform. The compatibilist thesis, on the other hand, although consisting of a mosaic of different views [46], ultimately just comes down to not caring about our behaviour being determined by physics [51], but caring about preserving, at all cost except rejecting physical causal closure, our familiar, intuitive perception of moral responsibility while deliberately overlooking the fact that this perception is founded on a tacit libertarian free will intuition.

Hard incompatibilism

Hard incompatibilists view the epiphenomenalist scenario as advantageous because it would allow us to suppress feelings of anger, indignation, hate etc. towards wrongdoers, since we would know that their behaviour was determined by factors outside of themselves [[46], Chapter 11]. However, this argument is not really convincing, as it is clearly only an imaginative stratagem, a contrivance, perhaps well-intended, for preserving one's peace of mind in the face of predicaments by selectively focusing only on the implications of epiphenomenalism on the culpability of other wrongdoers. Any worldview, even one that is horrible, will contain some elements that could provide an optimistic person with grounds for optimism by isolating certain aspects of that worldview from the rest of the context. In the present case, the argument deliberately ignores other consequences of epiphenomenalism, even for the wrongdoers themselves who turn out to be mere puppets, much degraded compared to our normal conception of a human being, even one who chooses to do evil. Another such consequence is that if epiphenomenalism is true then there is nothing wrong with focusing on its consequences on one's own culpability rather than on that of others. A legitimate strategy then would be for one to force him/herself to become as vicious, ruthless, depraved, and wicked as possible. Due to human nature, he/she may have to persevere in this effort and resist any natural feelings of mercy, sympathy, empathy, guilt, disgust or horror that they may feel when harming, deceiving and exploiting others, until they get over the hump and the internal resistance subsides. All that matters is that the evil actions are brought to completion, because once they are completed it is retrospectively thus proven that it could not have been otherwise, that this behaviour was dictated by the laws of physics. In hindsight, the perpetrator can calm his/her conscience, since the mere fact that the actions happened proves that it was these specific actions that were determined entirely by the laws of physics (possibly including contributions from random, quantum mechanical, physical events).

A hard incompatibilist may protest that if one chooses such a path of action then the blame is on them, not on

determinism; for determinism offers also the perspective of suppressing negative feelings towards others, and whether one focuses on the positive or negative aspects of a theory is their own choice. But such a protest makes the hard incompatibilist inconsistent with his/her own beliefs, and reveals that he/she does not even understand determinism. For, if determinism is true, then choice is just a figure of speech, an illusion, and whether one "chooses" to focus on the positive or negative aspects of a theory is entirely determined by physics, over which he/she has no control. Hence any person that "chooses" such a path of action, even if during the process they think and feel that they are making a choice, perhaps even forcing themselves to overcome innate sentimental resistances, was in fact programmed to will, feel, decide, resist, overcome and act in this way; they are just a physical mechanism, governed entirely by physics, albeit its functionality includes the production of feelings such as willing and deciding. So, if epiphenomenalism is true, then any person that actually "chooses" to adopt such a radical wicked path of action can, in retrospect, rest assured that it was not their fault but it was a consequence of the physical structure of the universe.

One need not adopt as radical a strategy as this. They could simply use a similar strategy to justify their laziness and inactivity, or to justify passivity towards evil and wrongdoing. For example, if a girl is being raped or someone is drowning in the sea, I could choose to do nothing about it. In retrospect, since I did nothing about it, this means that the physics of my body determined that I would do nothing about it; my body just happened to be built that way. Furthermore, I can remain completely psychologically passive and serene as these events are taking place, since the rapists are no more evil than the waves drowning the person, as both the rapists and the sea are physical systems behaving according to the laws of physics. So am I, whence physics determines that I remain passive[26]. And the girl raped and the person drowning are similarly just physical systems, and their agony, pain, and distress are due to the way their brain is structured; possibly, in the future we can design "better" brains such that such feelings do not arise, and people always feel happy, serene, and enjoying themselves whatever circumstances they are in, including drowning and being raped.

In fact, if we consider again the free will skeptic argument that such skepticism is beneficial because it allows us to suppress our desire for retribution against wrongdoers, it is clear that the same skepticism can be used to justify such desire for retribution. For suppose someone thinks that he is wronged and wants to take revenge; even if he feels that revenge falls short of the highest moral standards, he could nevertheless take his revenge appeasing his conscience by the thought that since the desire for revenge, as well as the revenge itself, occurred, they were physically determined. Libertarianism provides better means for suppressing one's vengefulness; it does not deny that external factors such as biology and the environment exert pressures on a person's will (without, however, making them a puppet) presenting to them a picture of reality somewhat different from that perceived by another person or causing strong desires that are hard to resist. So, it is not fair to morally compare two persons simply by evaluating their moral response in a particular situation, but a fair comparison would require that both persons had exactly the same biological background, upbringing, and life histories in general. Hence, in order to be fair when judging someone, one must put him/herself in the other's place first, imagining what it would be like to grow up in the same environment etc. Given that this could never be done with sufficient accuracy, it is best to soften one's feelings towards wrongdoers, thinking that there may be mitigating, or even exculpatory, circumstances associated with their choices and actions. But even if, given one's genetics and his/her history of environmental influences, his/her immoral behaviour seems unjustifiable, one should always keep in mind the possibility of repentance at some point in time, of a volitional change of mind, as experience with others and with our own

selves reveals. In my opinion the purpose of this life is for us to get acquainted with good and evil and their consequences, and then freely choose between the two. As long as we live, both options are open to us, and we do not know on which side each person will ultimately volitionally settle. And no person is entirely good or evil; oftentimes love for someone prevails so that we choose to forgive them even though their actions were not justifiable, and even without them having repented. Therefore, we have this ability, whereas under determinism "forgiveness" (which bears only a superficial semblance to actual forgiveness) follows logically and not lovingly since a person is no more responsible for his/her actions than a programmed robot or any other machine or object. So, this "libertarian" line of thinking can achieve similar, but not exactly the same, results as the hard incompatibilist line. The similarity lies in that both strategies can be used to tone down animosity towards wrongdoers. But the hard incompatibilist line is ultimately selfish, because the goal is to preserve one's own peace of mind, through indifference and apathy, since everything is determined, whereas the libertarian line involves sympathy, even for the wrongdoers, whose choice of evil over good deprives them of any real happiness in life, providing them with mere simulacra of happiness.

Hard incompatibilism mistakenly regards determinism as only eliminating a single evil (or at least it perceives it to be so) aspect of our world, when in reality it eliminates all active (volitional) good and evil by transforming them into natural good and evil, that is to fortuitous physical events and unfortunate physical accidents. This very elimination would be a very unfortunate circumstance, a natural evil. We would have no actual saying in the world, we could not actually influence it (including our own selves) in any way on our own. Yet we would retain the illusion of free will, and furthermore we would be forced to play along with this capricious, or rather sadistic, whim of nature: if we do not exert effort (actually, the illusion of effort) then our affairs in life will quickly turn south, despite our "effort" not actually having even the slightest impact on the world, since everything is actually driven by physics. In this perspective, a decision by one to turn completely evil, as a reaction or revenge against this whim, does not seem unjustified, and besides, it would only be natural evil that he turns to, while his “decision” would itself only be an evil event of the natural kind, an unfortunate accident. And in the deterministic world unfortunate accidents can only be avoided by fortuitous coincidences.

Hard incompatibilists downplay tremendously the consequences of determinism and think that it has only a minimal impact on the ethical, social, personal, and inter-personal-relationship aspects of life [[46], Section 11.8] — essentially, it only annuls a few notions like "basic desert", "moral responsibility" and "blameworthiness"[27], which (according to them) only played a marginal role anyway [47]. Their arguments for this seem to be based on an unconsciously libertarian view of persons masqueraded as deterministic, since they try to rescue those aspects of human life that determinism demolishes by referring to reasons that one may have to choose to behave similarly to what they would behave if libertarianism were true instead, in an effort to motivate one to choose a moral behaviour even under determinism. For example, Pereboom summarises it like this:

> *The free will skeptic would resist or disavow resentment and indignation, but she would not be exempt from disappointment, sorrow, and concern for the offender upon being wronged. She would have remorse for her own immoral actions grounded in sympathy with those affected, and with moral resolve she would take effective measures to eliminate dispositions to such actions, reconcile with those she has wronged, and to restore her relationships. When hurt by another, she might blame in the forward-looking sense ... but upon a commitment on*

> *the part of the other to eliminating the disposition to act this way, she would acknowledge this commitment and cease to regard the hurt as an obstacle to her relationship. She would be thankful and express joy toward others for the good things they provide for her. Her beliefs would pose no obstacle to love.* [[47], p. 192—193]

The above quote is full of explicit and implicit references to reasons and choice, which is the natural libertarian language but is completely misleading if used to describe a deterministic reality. All instances of the word "would" ("would resist", "would disavow", "would have remorse", "with moral resolve she would take effective measures", "she would acknowledge", "she would be thankful" etc.) should be replaced with "might": she might have remorse, if the molecules of her brain happened by chance, by contingent factors, to be in the right arrangement, if her neural circuitry happened to have a certain pattern, if the connectivity of certain neurons was a certain way, if the concentration of a certain substance in her blood was above a threshold etc.; otherwise, she would not have remorse. Whether or not these physical conditions hold is not up to her, but it all depends on fortuitous or unfortunate circumstances, such as maybe her diet, the properties of proteins, the DNA she inherited, her mother's lifestyle when she was pregnant, the solar radiation, and ultimately the meaningless and purposeless laws of physics and chemistry.

(In fact, whether "fortuitous" and "unfortunate" are even meaningful in determinism is debatable. These terms presuppose that our understanding of reality is genuine, that there are reasons why something is good or bad. But in determinism all our behaviour, including our mental understanding, is caused by factors that ultimately do not have to do with those reasons, but, say, with unrelated mathematical laws of physics. Hence it is likely that we are merely determined to see things as good or bad, with good and evil not existing objectively.)

Since reasons have no causal power (except in the weak and indirect sense of "motivation" in the case that libertarianism is true), in determinism they only seemingly appear to play a role in determining the behaviour of agents but in reality they have absolutely none. Determinism means that there are factors that cause the behaviour of agents, and reasons, by their nature, lack this capacity. The apparent role of reasons is an illusion which is caused by the same factors that cause also the behaviour; this is the essence of epiphenomenalism. The only case they could have an indirect role is if a third agent that possesses libertarian free will and who is motivated by these reasons (e.g. God) is involved in the determination, pulling the strings on the factors that determine the deterministic agent's behaviour; but even in this case it is misleading to say that the reasons determine the latter's behaviour in a language that implies that it is the understanding and contemplation of these reasons by the agent that leads them to choose to perform the determined action.

But the psychological, or "intentional stance" language used by hard determinists alludes to libertarian free will, which is necessary in order to provide a false unconscious, psychological assurance to the reader that some sort of secret libertarian free will and choice are preserved even under determinism. This is necessary in order to retain the psychological structure of the human world which would otherwise collapse, and hence inspired from Spinoza's quote from Section 5.5 we can say that human life feels meaningful and bearable only as long as one is not conscious of the causes that determine their feelings, thoughts, and actions, or those of others.
So, even hard determinists eliminate such thoughts of physical causality from their consciousness, limiting their contemplation of determinism to a theoretical sphere, and only practically applying it in a selected minuscule fraction of

human life. Our behaviour is portrayed even by hard determinists to be driven by feelings, judgements, reasons, morality etc. deliberately forgetting[28] that epiphenomenalism (including physical determinism) means that none of the psychological states of an agent have any causal power whatsoever, but both the psychology and the behaviour are caused by the physics of the body, by insentient laws of mathematical structure involving time, space, mass, charge, etc. Psychology is an effect and not a cause in the epiphenomenalist scenario, but the opposite is implied by hard incompatibilist quotes such as the above.

Compatibilism

Compatibilists, on the other hand, realise the severity of the consequences but try to avoid them by the arbitrary, ad hoc, unjustifiable, dogmatic and unsupported assertion that despite the physical determination of agents' behaviour, they are nevertheless morally responsible [52]. However, it should be clear that the aforementioned consequences are not cancelled even the slightest by the compatibilist thesis. These consequences depend only on determinism (including physical randomness), which both compatibilism and hard incompatibilism acknowledge. If I deliberately do something evil, both hard incompatibilists and compatibilists agree that it was the physics of my body that determined my actions; my body just happened to be configured in such a way (by chance, physics, etc. — factors all of which, ultimately, are completely out of my control) that in that particular situation I found myself into I would both want to perform the evil action and eventually perform it. The difference in opinion is only that hard incompatibilists recognise that if my behaviour is physically determined then it would not be fair to hold me morally responsible for my actions, whereas compatibilists contend that I should nevertheless be held morally responsible for my actions. The compatibilists' justification for why this should be so is lacking, but in the examples they put forth to support their thesis they typically appeal to the intuitive blameworthiness (or praiseworthiness) that we naturally associate with actions, overlooking that this intuition stems precisely from a worldview that intuitively assumes physical determinism to be false when it comes to persons and their will; furthermore, it overlooks that it is the same intuition that they appeal to which informs us that if someone's behaviour is physically determined and therefore out of his/her ultimate control then they are not responsible for that behaviour. So, the compatibilist thesis can be summarised as follows:

1. Determinism is true.
2. People are morally responsible.
3. Therefore, moral responsibility is compatible with determinism.

Premise 1. comes from belief in physicalism, and premise 2. is supported by appealing to intuition. Therefore, if both premises are accepted to be true, the conclusion 3. follows necessarily. With this stratagem, compatibilists attempt to avoid having to explain directly how physical determinism does not rule out moral responsibility despite their obvious antithesis. If one is fooled into believing that both premises 1. and 2. are true, then they have to accept that the apparent incompatibility between determinism and moral responsibility must be a misconception, since the truth of both is a fact[29]. The compatibilist thesis thus seems like a utilitarian psychological trick, a manipulation tool to exert pressure on people to abide by the moral rules, for the benefit of society. By sophisticated, albeit at bottom line meaningless arguments,

compatibilists try to convince people that they are morally responsible; the success of this endeavour depends on people not fully understanding these arguments, but being impressed or intimidated by their sophistication into believing the compatibilist thesis. This psychological procedure was just described in “intentional stance” language, according to the terminology of Dennett [24]. However, since compatibilism typically assumes physicalism, it is useful to describe it also from the more fundamental "physical stance" perspective. In this more accurate description, the preaching of compatibilist tenets is likely to cause physical changes in the listeners' or readers' brains (by the propagation of sound or light waves that will induce chemical processes in them) such that they will behave more morally. Progress in science and technology will allow us in the future to make these same changes in people's brains directly, by medical procedures, without the need for preaching, or even any thought process on behalf of the subjects: they will possibly undergo general anesthesia, and when they wake up they will be more moral, with their brains modified appropriately. In fact, we could make even better changes than those achievable by compatibilist preaching, since the latter is an indirect, empirical method and is performed without complete knowledge of the effects it will have on the subjects' brain — hence it only works on some people; direct brain modification, on the other hand, can offer precise control of the behaviour, down to the last detail provided that we have sufficient understanding of the brain structure and functionality. Hence, the characterisation of compatibilist preaching as a "manipulation tool" is befitting, as "manipulation" is the term commonly used for interfering with someone's brain in debates about the topic of free will. In typical fashion, compatibilists may contend that there is something fundamentally different between the two procedures of changing one's mind by preaching and by direct brain modification. However, according to physicalism, all the psychological procedures that our mind undergoes are reducible to the physics of the brain, so, as stated above, there is nothing fundamentally different between these two procedures. Preaching is, in essence, just a physical means of modifying one's brain (the accompanying mental processes are just epiphenomenal, having no causal power of their own). Not to mention the fact that compatibilist argumentation is pure manipulation, since the arguments lack substance.

In reality though, if one sees through the sophistry, the compatibilist arguments have no strength. Suppose that physical determinism is indeed true, and I am such a person as described above who has decided to unreservedly follow a path of evil. I know that my choice and determination to follow such a path are determined by the laws of physics, and after I have performed any evil action, the very happening of that action proves that it was physically determined and it couldn't have been otherwise. My will, decisions and actions were actually physical events that occurred according to the laws of Newton, Maxwell, quantum mechanics, etc., laws over which I have no control[30]. I know this, and the compatibilists know this, and we are in agreement on this issue. However, compatibilists go on to decide that I am morally responsible for my actions nonetheless. In view of the mechanistic origin of these actions through the laws of physics, this declaration by the compatibilists seems vacuous and lacking in substance. It has no theoretical value, yet it can have practical value if it succeeds in manipulating people into abiding by the moral rules. Nevertheless, I am not intimidated by the compatibilists' vacuous arguments and see clearly that they are completely irrelevant to the fact that my actions were physically determined, and to the fact that my future actions, whatever they may be, once done, will have been done so because they were also physically determined. Hence, these arguments have no force to make me swerve from my chosen strategy. Furthermore, I need not feel frustrated by the compatibilists' insistence despite the obvious truth. For I know that their thoughts, beliefs, desires, motivations, intentions, efforts, actions etc. are physically determined, just as mine; their

compatibilism derives from the contingent fact that their brains happen to be structured in a certain way, over which they do not, ultimately, have control.

6.2. Good and evil

However, it is at this point that some credit has to be given to compatibilists in their debate with hard incompatibilists. It should first be noted that ethics does not have a palpable basis like physics, where a theory can be tested experimentally and disproved by undeniable facts. Mathematics and logic also have more solid bases than ethics because, although their entities are not physical, they are closely connected with the observable, physical world. Mathematical entities such as numbers, vectors, lines, points etc. are idealisations or abstractions of actual physical entities or quantities; the truth of mathematical or logical claims can be tested and confirmed or disproven. It is not so with ethics; no experiment or logical proof process can show or prove that a moral claim is true or false. Morality and ethics, contrary to physics, mathematics and logic, do not have to do with the physical world but derive from the intrinsic value of persons, of minds. Ethics ultimately has to do exclusively with the realm of minds, which is not "observable" in the same way that the physical realm is, but is only observable in an introspective, private way that leaves room for subjectivity and different opinions between persons[31]. Nevertheless, everyday experience shows that there is common ground among people, a common core intuition, and hence we expect people, and consider them obliged, to meet some minimum moral standards, which everyone expects others to intuit the same way as he/she does[32]. The important question for our present discussion is this: is there an objective ethical standard which is independent of our perception of it, an independent moral reality, or are ethics and morality only subjective concepts, human conventions or inventions, or mere human illusions? Ethics and morality are founded on the reality of minds, and since the ontological status of minds is debated, the ontological status of morality and ethics is also controversial. If minds are substances and personhood is fundamental then the case for ethics being objective appears to be very strong. On the contrary, if physicalism is true then that case appears to be very weak. If minds are illusory phenomena, macroscopic manifestations of microscopic physical events, then our perception of ethics and our moral intuition seems equally illusory. If consciousness and mentality are reducible to physics, then it seems reasonable that ethics and morality should also be reducible to physics, that they should be ultimately derivable and explainable from physics. If physicalism is correct and everything is physical then good and evil, right and wrong, morality, are ultimately reducible to the apathetic, insensate mathematical laws that govern the physics of the universe, such as the laws of Newton, relativity, and quantum mechanics. If persons are mere complex objects that can be dissolved into their constituents, the fundamental physical particles that comprise them, and these can be recombined into other persons or into inanimate objects etc. then it seems reasonable that ethics is just an illusive product of our brain's chemistry. If our moral beliefs and our corresponding behaviour towards others can be fully physically explained with reference to biochemical mechanisms in our bodies, and the origin of these mechanisms can in turn be fully explained with reference to evolutionary processes driven by natural selection such that the reason that we exhibit such beliefs and behaviour is that they increase our chances of survival, then moral nihilism appears very likely, as the case for our moral beliefs being actually and objectively true weakens considerably.

Modern hard incompatibilism is founded on two pillars: physicalism and the objectivity of moral responsibility. Hard

incompatibilists would be right in their claim if both of their premises were true, but it seems unlikely that these premises are compatible with each other. In this respect, it is in fact compatibilism that seems more consistent with physicalism, as long as it acknowledges that moral responsibility does not have an objective status but it is a man-made concept, which we are thus free to define as we please. Compatibilists therefore define it in a way that mimics our false (assuming physicalism) intuitive conception of objective moral responsibility, in a consequentialist aim to benefit society, by preserving the social benefits that were gifted to us by evolution through this illusion. Although there are a couple of problems with such a compatibilist thesis, I think that it is more consistent and more compatible with physicalism than hard incompatibilism is. The first problem is that, as previously noted, the effectiveness of the compatibilist strategy depends on concealment from the general public of the truth of the very moral nihilism that lies at the foundations of the thesis. The second problem is that the thesis acknowledges that something is "better" for people and society than something else (namely, the upholding of moral rules, however man-made, is better than moral skepticism), but the ontological status of "better" in a physicalist setting does not seem to be much different than that of "more moral". Hence the compatibilist thesis is not entirely consistent either.

Furthermore, consideration of good and evil from an epistemological perspective weakens the hard incompatibilist thesis even further. Good and evil, if they exist objectively, are not reducible to physics; they do not follow from Newton's laws of motion, or from Maxwell's laws of electromagnetism, or from special or general relativity, or from quantum mechanics. So, even if we accept good and evil as existing objectively, beyond the realm of physics, still they would be epistemically inaccessible to us if we are what physicalism assumes us to be. Good and evil cannot engage in physical interaction; they do not possess mass or charge, they do not attract or repel mass or exert any kind of force that could have even the slightest effect on the particles comprising our brains, they do not emit electromagnetic radiation etc. Therefore, there is no physical mechanism with which they could influence our brains in any way. Our brains function according to physical laws such as those just mentioned, none of which involves good or evil either directly or indirectly. Neural signals are generated and transmitted via chemical potentials, which have nothing to do with ethics — they belong to completely different, impervious realms. Good and evil or right and wrong cannot cause a neuron to fire, or a molecule or ion to move. Hence there is no evolutionary path which can lead to our awareness of the real, objective good and evil, if these exist, since that would require that they cause physical events in our brains, that they physically interact with it. Evolution by natural selection, in a physicalist world, works based on how physical changes in an organism's "design" (DNA) change its probability of survival in a physical environment that interacts with it in certain physical ways. Hence even if evolutionary processes resulted in our development of a sense of good and evil, this has a physical origin and helps us to survive by, say, organising into societies and helping each other. That our sense of good and evil or right and wrong refer to objective, non-physical realities would be illusory, while the real good and evil, even if they exist, would be epistemically inaccessible to us. If this is so, then the hard incompatibilists' argument fails, because it assumes that their intuitive conception of moral responsibility as objective is genuine, whereas in reality it is impossible to know whether such moral responsibility exists, and even if it does it would be something different than what we perceive.

It therefore seems that physicalism implies moral nihilism. Despite the fact that we think that we are aware of good and evil, and right and wrong, what we really perceive is effects of chemical reactions in our brains. In fact, it seems perfectly

possible, if physicalism is true, to restructure our brains such that our perception of good and evil is inverted; for example, we could see murder as a good thing, and charity as evil. And in the unlikely case that this turned out to be impossible to do, the impossibility would be entirely due to physical reasons, i.e. the physical laws would not allow something to happen in the brain, such as a violation of the energy conservation principle. Hence this very impossibility would confirm that good and evil are reducible to physics. In other words, when it comes to ethics, physicalism is a good fit with antirealism but not with realism [53]. What follows is that moral nihilism is an undesirable (or perhaps desirable for some) consequence of physicalism, and of epiphenomenalism in general.

6.3. Personal value and equality

In the modern era, a strong aversion towards racism and sexism and their practical manifestations, racial and gender discrimination, features in the worldviews prevalent in developed societies. And rightfully so, in my opinion. On the other hand, physicalism is another main ingredient of these worldviews; it is the theory that everything ultimately comes down to physics, which therefore subordinates ethical principles, such as the aforementioned racial and gender equality, to corollaries of physical principles. But if a person is nothing more than the complex physical arrangement of his/her body, and there are all sorts of bodies in nature no two of them being exactly the same — in fact this diversity lies at the basis of the very formation of living organisms, through evolution by natural selection — then it seems that it is *inequality* rather than equality that is more congenial to physicalism. So, is physicalism compatible with these other aspects of the modern Western worldview, or is the latter inconsistent?

The lazy but incorrect answer is that, being parts of the same worldview, they are certainly compatible. But human psychology is not always rational, and various motivations, purposes, circumstances and the desire of people to be part of a group even if that requires turning a blind eye towards the inconsistencies of its ideology can, and do, result in incoherent ideologies. For example, past and modern "conservative" worldviews often bundle together Christianity with theses that are irrelevant or even incompatible with it such as capitalism, nationalism, militarism, the right to bear firearms, even with racism and polygamy in extreme cases. Physicalism and determinism are also incompatible with the teachings of Jesus, yet accepted by some Christian communities who are blind to the conflict.

It seems that an analogous situation holds for the modern "progressive" worldviews. In particular, in a physicalist interpretation of reality it is not possible to find solid grounds for establishing a normative status for racial and gender equality, or equality between persons in general, however desirable this is and despite being considered as one of the pillars of these worldviews. A physicalist is forced to accept that a water molecule, a virus particle, an amoeba, a mosquito, a frog, a mouse, a dog, a chimpanzee and a human are hugely different in terms of intelligence, cognition, consciousness, and mental capacities in general. They also plausibly differ in value (whatever this may mean in physicalism): a human life, for example, is far more valuable than that of a mosquito. What is the reason behind these differences between the various organisms? In physicalism, there can be only one answer to this question. Since the physical structure of the body is all there really is, any comparison between organisms ultimately comes down to a comparison between the functionalities of their bodies; some species have "designs" that allow them to perform more functions, or perform the functions better, than others. When it comes to valuing organisms, judging one to be superior to another, we typically focus on their mental/intellectual/cognitional capacities which (according to physicalism) are products

of their brains' "design", which gives rise to the corresponding functionalities; hence bears and tigers, for example, are regarded as inferior to humans although physically stronger.

The potential of each organism is encoded in its DNA, and the degree of actualisation of this potential is a function of the environmental influences during its lifetime. This allows at least two different bases for comparison. One possibility is to compare the actual physical structures of the organisms, which resulted from the combined effect of the DNA and environmental influences. The other is to compare only their designs, their DNA, which removes the environmental influences; it may be that, potentially, one organism could have developed so as to have more abilities than another (because its DNA is better), but due to unlucky circumstances its environment held it back. Of course, when comparing organisms across different species it is most likely that both definitions will give the same result, as the DNA will be the dominant factor. On the other hand, when comparing individuals of the same species, the environmental factors can be equally or more important than the genetic factors, since the latter will be very similar between the two individuals. Nevertheless, genetics certainly play a role; if we focus on humans, we know from everyday life that one is better at mathematics, another has a musical talent, another has a rich imagination, someone else has a very strong memory etc., and these differences can not, in most cases, be attributed entirely to the environment, similarly to physical traits such as height, strength, speed, endurance, physical appearance etc.

So, if a person is identical to his/her body, if the constituents of the body is all there really is as a substance, then it follows that a comparison between individuals, concerning any aspect of them, merely comes down to a physical comparison between their bodies. In particular, in the case that the comparison concerns intellectual and mental capacities, the comparison would mostly come down to a comparison between their brains. And since, in practice, two bodies/brains are never identical, it is natural that one body's design is superior to that of another in certain respects. In fact, the notions of superiority and inequality lie at the very core of the mechanism of evolution by natural selection, where the superior designs (in terms of survival in a certain environment) are "selected" for survival. The basis of evolution by natural selection is the genetic diversity between individuals and the diversity in behaviour, abilities, traits etc. that it causes. The superiority of some designs over others is evident if, as noted, we compare a mosquito to a human. Of course, one may argue that there is a huge difference between comparing a mosquito to a human and comparing two humans. This is undeniable; nevertheless, if physicalism is true, then this difference is quantitative, not qualitative: the differences between any two humans are minuscule compared to those between a human and a mosquito, but they are the same sort of differences. The notion of species is a human convention, despite its great practical value. Different species do not refer to different kinds of substances. Humans and mosquitoes are both, from a physicalistic perspective, complex physical objects built of the same kind of biological technology. What is objective about species is that genetic differences within them are smaller than those across them, but where to place the dividing lines is necessarily somewhat arbitrary, because categorisation into species is an artificial discretisation of something that is essentially continuous.

The important question concerns the status of the claim that all persons are equal in some important sense, e.g. in terms of value. There are three possibilities to consider: either this claim is objectively true due to the ontology of persons (the nature of persons is such that they are necessarily all equal), or it is contingently true (the nature of persons does not

entail equality, but due to contingent circumstances they happen to be all equal at present), or it is false. Intuitively, I think that most people (including myself) believe and desire the first scenario, that persons *must* be equal — that there is a metaphysical necessity, related to the ontology of persons. However, this is the only option that is incompatible with physicalism. This is evident if one considers that in physicalism any person is ultimately reducible to just his/her body, or if one tries to describe a person using "physical stance" language rather than "intentional stance" language. There is nothing necessitating that one physical system is equal to another in any sense (except in that they are all physical; but in this sense we would be equal to a rock, a chair, a glass of water etc.). Since the bodies differ, the persons differ correspondingly. And it is most likely that one body will perform certain functions better than another because its design or layout happens to be better. In physicalism, a person is just a physical machine and since the blueprints of the physical machines (bodies) differ, the same holds necessarily for the persons which are ultimately identical to, or functions of, these physical machines. The best that physicalism can offer is an equality of the second kind, a contingent, coincidental one. Such, for example, is the best achievable by the argument that people are equal because even though one's performance in some activity (e.g. mathematical reasoning) may be below average, he/she will excel in some other area (e.g. they may have a talent for music). Obviously, there is nothing necessitating that one's weaknesses are completely balanced out by their strengths so as to achieve exactly the same "score" (in any scale of abilities) as all others. Even if this turned out to be true it would be by some incredible coincidence and not by logical necessity. Similarly, contingent truth is the best achievable by arguments that attribute people's weaknesses entirely to the environment in which they grew up rather than to their genetics. Essentially these arguments claim that, despite appearances, the genetics of people *happen* to be equivalent. So, at best, under physicalism all persons could turn out to be equal *by some incredibly lucky coincidence*, but most likely they are not.

From this core problem, other derivative problems arise such as the equality between races and genders. From a philosophical point of view, these are more artificial because they refer to differences between statistical averages of different populations and not between specific, real individuals (a comparison between the statistical averages of the races or genders of two individuals does not determine the outcome of the comparison between the individuals themselves, and therefore is only of theoretical significance). Nevertheless, in real life racism and sexism have often been the source of severe injustice and hence modern developed societies, particularly multiracial ones, regard such views as particularly heinous. Again, physicalism can offer no normative answer as to why races and genders should be equal. In physicalism, all the potential value of a person lies in his/her genetic blueprint (the actual value derives also from environmental influences) and therefore the only sensible definitions of racial or gender equality would be on the basis of comparisons of suitable statistical averages of genetics across races and genders. If these averages happen to score equally on some scale, this would be due to some wild coincidence. Inequality, however slight, is the expected result according to the principles of biology and evolution — inequality is, after all, the driving force behind evolution. Currently, the problem is masked by the slightness of the differences between humans, which one can therefore attribute to environmental factors to avoid having to face it. But, what if in the future, when our understanding of genetics and our technology have progressed significantly, we become capable of designing and constructing organisms ourselves, rather than leaving it up to evolution? This seems like a very likely scenario. For example, wealthy people may then have the opportunity of hiring geneticists to design their offspring's DNA sequences according to desired traits and abilities that

they want their offspring to have. In a physicalistic world, since a person is reducible to his/her body, and the complexity of the latter has no upper limit, there is no reason why, just as a human being is greatly superior to a mosquito, there could not be designs that would be as superior to a current human as the latter is to a mosquito. Suppose a wealthy individual is born as a regular human, but he pays his way into transforming himself into a super-human by genetically modifying his DNA, possibly also using implants etc. How would he then compare against regular humans in terms of value? How would he compare against his previous self?

Should he even be considered to be the same person as before? (Of course, in physicalism sameness of self is a problematic concept because persons are merely composite objects, just like the ship of Theseus). In such a world, would equality be meaningful? In what sense? What would be the physicalist basis of it?

Essentially an implication of any theory that regards a person as a composite entity rather than a substance is that the person has no absolute, objective intrinsic value, but the real value lies with its constituents and what it is about them that gives rise to the person (their arrangement, properties etc.). The person is nothing more that a perceived, subjective, conventional entity like the composite objects of Comment 7[33].

Physicalism, in particular, considers a person to be reducible to a complex physical system and hence any physicalistic definition of value of a person necessarily comes down to a metric of physical characteristics, to a physical evaluation of that system. Our intuition that a person's value lies in its psychological characteristics, that they live, feel, will, think, choose, exist, etc. is as illusive as these apparent characteristics. They appear to us this way, but in reality they are just physical processes in the brain. The value of a person is as illusive as the existence of the person itself; a person has value only in a relative, subjective sense. Thus, modern humanistic worldviews are internally inconsistent because they are founded on the mutually exclusive principles of human value and physicalism.

Similarly, panpsychism also postulates an illusive, relative and subjective existence of macroscopic persons, viewing them as composite objects as well, as collections of physical particles, whose properties are determined by the structure and properties of their constituents; the only difference with physicalism is that these particles are assumed to have an extra, non-physical property, (micro)consciousness. In that case, what has real, objective value would be the (micro)consciousness of the fundamental particles, whereas the composite consciousness of the human person would still be subjective and illusive and of no real, objective value. Destruction (death) of the human person would have no implication on the underlying particles and their consciousness, which could then recombine and form other macroscopic persons or objects; death would, from an objective point of view, be a mere rearrangement of the conscious particles, involving no actual loss, with everything that really exists being preserved.

Consider also philosophical theories that, irrespective of what they consider to be the objective constituents of a person, even from the subjective point of view consider a person to be just a bundle of mental states or the collection of the "peripheral qualities" in the terminology used in the present paper, without an underlying substance. Obviously, such theories also cannot justify assigning any objective, absolute value to a person. In contrast, in substance dualism a person exists objectively as a substance and has an objective and absolute value. What is of absolute value is the unique, singular, irreplicable first-person, alive, point of view of each particular person, which cannot be mapped onto any physical pattern, or combination of constituents in general whether physical or not. Personhood, essentially the origin point of each

particular first-person perspective, is an "all-or-nothing" thing, and there lies all the value of a person. Hence all persons are of equal yet unfathomable value[34].

In contrast, the "peripheral qualities" of a person, in the terminology I used previously, such as intelligence, memory, knowledge, temperament etc., and related mental events, vary throughout that person's life, and are likely, to some extent, correlated with physical structures and events in the brain, respectively. (Nevertheless, even these are unlikely to be entirely correlated to the physics of the brain, as argued in Sections 3 and 5).

I am the same person — and hence have exactly the same intrinsic value — now in my prime, as I was when I was a helpless fetus[35] or baby, and as I will be when I grow old and lose my powers, or when I am asleep and most of my powers are deactivated. The same will hold if I somehow, using science and technology or by whatever means, transform into a super-human. The core of my existence and my value are independent of the particular powers and abilities that I possess at any given time. Having greater intelligence, memory, etc. does not make me more inherently valuable in the same way that driving an expensive car does not make me superior compared to someone who drives a cheap one. Intelligence and other powers can be seen as add-ons, which are to a large degree due to chance (we don't control what traits we inherit) but are also cultivated, and possibly, in the future, engineered. But what is irreproducible is the self, the unique identity, the first-person center. The inherent value of each person is therefore the same. There is another kind of value, not easy to define, which depends on a specific aspect of personhood: free will, and in particular if one chooses good or evil. That is up to the person themselves. In modern scientifically-minded societies, what is often admired is people's scientific achievements, which depend on intellect (in lower intellectual strata even bodily characteristics such as beauty or athletic ability are also admired first). However, intellectual ability is also a matter of chance, and we all know that it is not correlated with ethical character. It is the latter that should be admired.

If personhood is an "all-or-nothing" thing, then are virus particles, blood cells, insects, frogs, dogs etc. persons? And if so, are they all of the same value as us? Well, in my opinion, if they are persons, yes. If an ant has a first-person perspective, then the fact that it is unintelligent compared to us is contingent, related to the body it happens to have, but is not part of its nature as a person; by augmenting its nervous system to include a human-sized brain it would acquire all the intellectual abilities that we have, from the same first-person perspective that it had in the first place, without becoming a different person. But I do not think that all organisms are persons; since personhood does not follow logically from physical construction, all organisms should, a priory, be just biological machines, without consciousness. It is another factor (God in my opinion), unrelated to physics, that brings about the mind-body union. Hence this factor (God) may decide which bodies will be joined with minds and which not. If I had to guess, then like Descartes I would guess that only humans are persons while the rest of the animals are biological robots / philosophical zombies. However, this is impossible to know due to privateness of each mind, and hence I regard the contemporary movement in favour of the welfare of animals as justified.

7. Summary and final thoughts

Modern philosophy of mind is characterised by continuous but failed efforts to devalue the mind, to demystify it by reducing it to more mundane constitutive elements. To me, this fixation of philosophers and scientists with deconstructing the mind, disproving its fundamentality and finding out what "is behind it" is depressing and bewildering, especially since they themselves are persons, minds. Hopefully, the preceding sections provided compelling reasons to search in the opposite direction, attributing to the mind a fundamental rather than derivative status and role in reality. In my opinion, no matter how fascinating the mysteries of the physical world are, they pale compared to the mystery of the mind, while the substances of the physical realm also seem completely devoid of value when compared to living minds.

Conviction in the reducibility of the mind to simple constituents leads to the concomitant belief that our first-person perception of it is illusory, similar to our perception of macroscopic physical phenomena, and that therefore in order to get to the truth about the mind we should distrust first-person observations and rely on third-person, scientific observations instead. Due to the nature of the mind this approach is doomed to failure as the first-person perspective which is the very object of investigation is inaccessible to outside, third-person means. Science, as a practice that admits only third-person observation, has severe limitations in this case and can only offer help by examining aspects of the physical reality that have a contingent correlation with mental phenomena. On the contrary, any attempt to understand the mind must begin with an introspective examination of it, from the first-person perspective. Furthermore, what we call third-person observations are really our first-person experiences of things outside of us. Hence if the first-person point of view is illusory, then we cannot trust even third-person observations. Introspective study of our own selves can reveal to us many interesting facts about reality, not revealable through the third-person methods of science. This essay has made an effort to highlight some of the deep aspects of personhood.

In Section 2 it was argued that first-person experiences are not derivable from physics — either known physics or that awaiting to be discovered — because the quantities that physics deals with, such as mass, charge, length, time, force, etc. are completely alien to the nature of conscious experience, as is any combination of them. Any law that could count as "physical" would involve these physical quantities and their combinations and derivatives, and nothing appearing in the mathematical equations that express it could logically imply the emergence of consciousness. Hence consciousness is not deducible from, or reducible to, physics.

Furthermore, in Section 3 it was argued that minds have the metaphysical capacity for genuine understanding of reality, which is something that is impossible to reduce to physics. That is, metaphysics is a view of physics from the outside, from above, and is not explainable or deducible from within the realm of physics itself, it is not analysable or reducible to the simple physical laws that it explains. Solving the equations that express Newton's laws or the laws of quantum mechanics or relativity will not produce an explanation of why mental, conscious understanding arises. It will merely explain why this or that ion in our brain moved in that direction or why there is an electrical potential across a neuron's membrane, which are the kind of entities and interactions that physics is about.

Section 4 is arguably the most important part of this essay, and explains that mental phenomena, which are the focal point of the philosophy of mind, are themselves of secondary status and importance compared to their carrier, the person. It was shown that the person is not reducible to any combination of constituents, because its most important aspect, its identity, who that person is, can bear no (a priori) special relation to any combination of outside elements compared to any other (existing or potential) person, that would explain why this particular combination, and not some other, gives rise to

that person, and why it gives rise to that person and not to some other. All persons are symmetrical, exactly similar, from the point of view of any outsider; the core of each person, its identity, *who* he/she is, is visible, palpable, perceivable only from that person's own perspective, only from within its own self. Furthermore, any combination pattern could be duplicated, whereas a person is non-duplicable (the same person cannot exist in two instances). Hence persons are simple, i.e. non-divisible substances; in fact it seems to me that the person/self is the quintessential simple substance, in the sense that its simplicity is metaphysically (logically) necessary, whereas the simplicity found in physics (e.g. fundamental particles) seems to be only physically necessary, contingent, it is only an empirical fact. Of course, these arguments can be understood only if one makes a serious effort to introspectively examine his/ger own self, and Section 4 attempted to offer aid for this process through several thought experiments.

If then it is accepted that for each existing person there was a time when they did not exist, the question arises of how that person, completely inaccessible from the outside, could be brought into existence from pre-existing outside factors that completely lack access to it. This "creation problem", the logical impossibility of creation of a particular person realised by outside factors, is a hard problem not only for physicalism but for all theories of the mind, dualism included. It forces us to become sceptical towards the validity of human logic, to recognise that we have limitations in our powers of understanding reality, and to be more reserved in our confidence when judging things to be impossible[36]. I then expressed my opinion that the source of reality is a Mind, God, who, in a mysterious way, has direct access to each created person and selected them even before they existed. My belief that the foundation and source of reality is personal (a Mind) rather than impersonal is based on my perception of a chasm in value between the personal and the impersonal. The physical seems to me to be insignificant compared to the mental, in fact subservient to it; it is from the mental that it acquires its relative value, even its relative existence. As I've mentioned, I find that even a single, finite, human person is infinitely more valuable than all of the vast impersonal, inanimate universe. Hence the structure of reality must have mental foundations, rather than the mental being a product of impersonal foundations.

Of course, this is not proof of the existence of God; if God wants us to be free to believe in him or not (as I think he does) then the world is designed in such a way that there cannot be any definite indisputable logical arguments for his existence. Believing in God should require free choice, and if the existence of God follows necessarily from logic then that choice is not free. Hence the structure of the present reality is such that it acts as a veil behind which God hides himself, or, using the background of Section 5.2, it is such that there are reasons both for and against believing in his existence. Furthermore, not all reasons for believing in God are of equal value, in both a judicious and an ethical sense — in fact some may be worse than certain reasons for *not* believing in him. In my opinion, the most fundamental reason should be the aforementioned acknowledgement of the singular value of persons, of the miracle of being alive. This should be a compelling reason for one to believe not just that God exists, but that he also values life/persons above all, and hence his character is that conveyed to us by Jesus — it is that of a Father and not a mere Creator.

The "creation problem" shows that the omnipotence of God should be taken more seriously and literally; of course, most theist philosophers acknowledge that God is "omnipotent" but seem to regard this as a figure of speech that means merely that God is more powerful than anyone else, or that he is as powerful as possible, as powerful as the rules of logic allow him to be. In other words, it is tacitly assumed that there is a set of rules or laws that are independent of him and do not

derive from him. Therefore, there is a reality, with its own rules and laws, that transcends God, that is conceptually prior to God, in which he has to exist. Hence the dubious notion of possible and impossible worlds, the latter being those that are impossible even for God to create. But the question then arises of where this fundamental, God-transcending reality originates from. It is larger, stronger, more primitive and fundamental than God, and is impersonal, unconscious, non-mental. Therefore, those who accept this view believe in a physicalism of sorts. The existence of such a God-transcending reality which is nevertheless graspable by the human intellect, is convenient for the philosophy of religion because it makes God amenable to human reason, it makes him explainable, understandable, analysable, just like everything else in the universe.

Similarly, God is usually understood as experiencing life in almost the same way as a human person: he empathises with persons by imagining what it's like to be them [54][55], he can be thinking of only one thing at a time[39], he experiences the flow of time just like we do [[37], §27.6], his omniscience is the knowledge of all true propositions. Perhaps this view is partially motivated by the idea that the only alternative to God being a person exactly like us is for him to be an impersonal force. But in my opinion just as we, the finite persons, are infinitely superior to an impersonal force, so God, the prototypical Person, is unfathomably superior to us; we are just shadows of the Personhood of God. Furthermore, even for us, I find it probable that our present form of existence and of experiencing life is transient, and that higher forms of existence that are presently inconceivable to us are in store for us in the afterlife[37]. Therefore, I find it very likely that the limitations placed on God by philosophers are untrue not only for God but, in some cases at least, even for finite, created persons in their fulfilled, eschatological state.

In Section 3 it was also argued that the correlations between mental phenomena/thought and physical structures and events in the brain cannot be metaphysically necessary but are contingent. Therefore, the psycho-physical laws that describe them are primitive, and we cannot hope to explain them. If I had to guess though, I would expect some kind of “functionalism” to be at play, where a mental phenomenon is correlated with some brain process that is related to a bodily operation whose function is somehow related to that mental phenomenon. Of course, "function" is a mental, intentional term, that expresses a notion nowhere to be found in physics, as argued in Section 3; therefore the basis for the correlation is ultimately mental and not physical — it is a functionalism based on mentality rather than physics. Nevertheless, I do not think that these correlations can be complete, i.e. I would expect that there are aspects of our thoughts that are not correlated to any physical processes in our brains. After all, the realm of thought seems to be continuous, whereas the brain is a discrete system consisting of a finite number of components. And, furthermore, whilst what I called "peripheral qualities" of the mind seem to exhibit a strong dependence on the physical aspects of the body/brain, the core of the mind/ego, its identity, who someone is, can exhibit absolutely no such dependency (Section 4). The psycho-physical laws are utilised by the structure of our bodies so as to turn external physical influences that our bodies receive from the environment into appropriate mental depictions of the physical world around us. This does not mean that how we perceive the physical world is how it actually is; the physical world does not actually have colour, sound, odour, is not cute, beautiful, ugly, frightening, funny etc. — the ability to experience all of these is a property of minds, and they are not properties of the material objects themselves. The structure of our bodies and the psycho-physical laws are such that they result in a nice, useful and functional mapping of aspects of the physical world onto mental representations that allows us to navigate in it, understand it, and live in it — but it is not the only one possible; by

re-wiring our brains we could perceive the world very differently (as perhaps bats experience it [12]), either by remapping our current experiences to different aspects of the physical world, or by having new experiences altogether (if our minds have the capacity for experiences as yet unknown which, however, are not triggered by the current design of our brains).

The current, functional and efficient mapping of our mental qualia to the aspects of the physical world around us could have been achieved by a process of biological evolution by natural selection, provided that epiphenomenalism is false and the mind has causal powers over the body. This would mean that different mappings would present the world differently to embodied minds and this in turn would have an impact on their decisions and their bodily behaviour (via agent-causation), which would potentially make some mappings more conducive to survival than others. The surviving organisms would then pass down these mappings, encoded in the DNA blueprint of their brains, to future generations. However, it should be noted that such an evolutionary process does not, and can not, explain how the psycho-physical laws themselves arise (evolution by natural selection is a means of producing efficient physical mechanisms to achieve some function, but that function must be physically achievable in the first place, and the psycho-physical laws are not physically achievable but are fundamental). Rather, given that there *are* such primitive, inexplicable laws, evolution can exploit them to match qualia with the physical world efficiently and advantageously by trying out different body/brain designs. Evolution is not the only way to achieve such a mapping — intelligent design is another; for example, if we knew the psycho-physical laws then we could ourselves design brains that make their bearers perceive reality in various ways. The problem in this case is, of course, that the psycho-physical laws are not completely knowable since we can not a priori know the full range of experiences a mind can have but each of us only knows the particular experiences he/she has had to date, as argued in [12].

But is this the way that evolution worked to produce our bodies? I am inclined to answer no, because this would imply that organisms have minds/consciousness since millions of years ago, and therefore that organisms as simple as insects or even simpler are ensouled, which I find unlikely. Rather, it seems more plausible to me that evolution occurred through purely physical mechanisms, involving nothing mental, and once it produced bodies of sufficient complexity God infused these bodies with minds and enacted the psycho-physical laws on these particular mind-body composites.

So, there are the two separate, fundamentally different and in principle unrelated worlds of physics and of consciousness, that are nevertheless linked together by the mysterious psycho-physical laws. In principle, we could exist and have mental experiences without a body; and the material world could exist without consciousness (other than that of God).

Concerning the physical world, my intuition is that in the grand scheme of things, eschatologically, when it has fulfilled its role, it will come to an end. Concerning the world of consciousness and the objects that exist therein, Plato would argue that they, the ideas and forms, the qualia and feelings, the meanings and concepts, are of timeless value and are the true foundations of reality. However, personally I am skeptical about that. What seems to be of intrinsic value is the persons/minds that have these conscious experiences, and not the experiences themselves. Thus, I find it likely that many of the elements of our consciousness are also transient, pertaining to this transitory reality, and will cease to exist once their role and purpose is fulfilled, giving way to a higher form of consciousness, with our experience of reality in the world to come being something that is inconceivable in our present form.

Persons are much more of an unfathomable mystery than is commonly thought, and this is a reason for great optimism.

We have real existence compared to the relative existence of the physical world, and therein lies our value, deriving from our likeness to the Existence that is God, the source of all reality.

Footnotes

[1] To be fair, such an accusation does not pertain only to physicalism, but most belief systems (counting physicalism as one of them), including the theistic ones that are in a hostile relationship with physicalism, contain inconsistencies that have lingered around for a long time so as to become incorporated into their canon. Such inconsistencies can seem obvious for outsiders, but insiders may avoid contemplating them for fear of shaking the foundations of their own faith or of losing favour within their community, or they may trust that the champions of their faiths, e.g. scientists in physicalism or charismatic individuals in religious systems, have the answers.

[2] In my opinion, an opinion shared by William Hasker as expressed in[56], this is the main argument in favour of physicalism over dualism. However, physicalists do not usually explicitly base their case on this argument. This is probably because the argument fails when one attempts to explain the correlations in terms of physics, because he/she immediately runs into the first two hard problems that will be discussed in this paper. Instead, physicalist treatises about the philosophy of mind usually very summarily and conveniently dismiss substance dualism right at the start by an argument against interactionism, that is put forth as a sort of "hard problem" for dualism. In particular, it is asserted that it would be impossible for an immaterial substance to interact with a material substance due to their different natures, or that any such interaction is precluded by the causal closure of the physical laws. Given that the physical-physical interactions that physics describes are themselves primitive and inexplicable at the most fundamental level, such a claim is bold and unjustifiable. The physicalist Jaegwon Kim [57] recognises the weakness of such arguments, and admits that they are "pretty much all we get from Descartes' critics and commentators". Then he proceeds to formulate his own argument that he believes to be more precise than the previous ones, according to which it is the lack of spatial location of a hypothetical soul that would make it impossible for it to interact with a body (or even with another soul), as interaction requires a spatial relationship (this requirement is obviously a physicalistic premise). This argument (discussed briefly in Sec. 4.1.) does not seem to be any better than the previous ones. Clearly, all of these arguments are question-begging and ultimately come down to this structure:

- Physicalism is true.
- Dualism is incompatible with physicalism.
- Therefore, dualism is false.

[3] Nevertheless, they still persist among some religious circles, in the form of "hylomorphism" or "Thomistic dualism" (which actually seems more like a kind of materialism) [7][58]. I find that they almost completely miss the point about what a person, an ego, is all about, describing the soul entirely from a third-person perspective, as we do with any physical thing, as something impersonal, a sort of "energy" or even "information" [[37], p. 267—270]. Apart from their aforementioned

vitalism, another distinctive feature of theirs is that they view any object, including humans, as consisting of matter and form ("hylē" and "morphē", respectively, in Greek — hence "hylomorphism"). For example, a statue is more than just the matter it is made of; what makes it a statue is its shape which resembles that of a person's body, and the consequent potential to generate a thought of that person in the minds of those who see it. "Form" is a broad term that can include structure, geometry, behaviour, functionality, utility, etc. In my opinion, which will be briefly expressed in comment [7], all these features that make the "form", that bestow to a complex object the significance and meaning it has to us (functionality, utility etc.), are relative to human minds, and complex objects do not have an independent existence but exist only relative to the minds that perceive them. For example, a statue may be a meaningless lump of matter for an alien who is completely oblivious of the human physical form, and the "form" of a female dog as perceived by a male dog is different from that perceived by me, who can't even easily discern between a female dog and a male one — in fact, the form of a male human is slightly different as viewed by men and women, and similarly for the form of a female human. Since objects have only a relative existence, one is free to define them however he/she pleases, and using the hylomorphic system is a legitimate choice. There is no right and wrong. When comparing theories about the nature of reality, we must distinguish between essential differences, and differences merely in definitions, conventions etc. which do not ultimately imply any actual difference. Cartesian dualism and physicalism are really different, because the former recognises two substances that can exist independently of one another, while the latter denies this. But if we compare hylomorphism and physicalism, it is not evident that they are incompatible. Sure, they use very different vocabulary, definitions, conventions etc., but it seems that in principle they could be different descriptions of the same reality. Hylomorphists claim that their definition of a human as matter and form (they use the term "soul" synonymously with "form" in the case of living organisms such as plants, animals and humans) somehow makes their human more than the physicalist conception of a human. But physicalists also do not claim that a human is just the matter of the human body. If we take a human and convert him into a pulp using a mincing machine, this would not satisfy the physicalist definition of a human being. In order to count as a human, the matter must have a certain configuration which, given the physical laws of nature, results in the complex functionality and behaviour exhibited by humans. If physicalism is true, we can still speak of "matter" and "form" in humans without being in error. What might differentiate hylomorphism from physicalism is the former's vitalistic claims, but even these could be construed in a physicalistic manner. If the "form" of the human is basically his/her physical arrangement and the ensuing functionality, then the laws of physics imply that this form causes the body to grow, sustain itself, sense, move, respond to stimuli, and even (if physicalism is true) to exhibit consciousness, thought, reason, etc. Given that hylomorphists remain at such an abstract and vague level of description and avoid any hard questions about the nature of the person/self, their framework seems compatible with physicalism. The only incompatibility would be their tenet that the "soul" survives biological death, a claim that seems inconsistent with the rest of their theory, like a foreign body patched into the theory just to make it conform to the ancient Christian tenet of the immortality of the soul.

[4] It is the weakest of the three if "conceivability" is taken in a weak sense as "imaginability", as discussed in Section 1, which is the sense that appears to be meant by proponents of the argument, since physicalists have responded that although we can imagine such beings as philosophical zombies, their existence is metaphysically impossible: if a being had exactly the same physical structure as us, then it would necessarily follow that it would also exhibit conscious

perception, thoughts, emotions etc. just like us — it is impossible for there to be beings exactly physically similar to us but mentally dissimilar. However, the aim of the present Section is to explain why, if a human is just a physical system that is completely describable by the laws of physics without leaving anything out, then it is metaphysically impossible that a human be anything more than a philosophical zombie. The fact that we are not philosophical zombies then means that the laws of physics are not sufficient to produce us, and that mental phenomena transcend the realm of physics. If the "zombie argument" is taken in this sense, then it becomes the strongest of the arguments pertaining to this hard problem.

[5] *Emergentism*, in its strong version, asserts that systems of physical elements have physical properties that are not deducible from the properties and arrangement of their constituents [59]. This view is widely rejected by modern science. For example, Broad [59] thought that the properties of oxygen and hydrogen tell us nothing about the properties of water, which is clearly not true, as long as we are referring to the oxygen and hydrogen *atoms*, which are the constituents of a water molecule. There is also the influential paper by Fodor [60] which claims that the "special sciences", i.e. all sciences other than fundamental physics, are not "reducible" to fundamental physics. However, this is not intended to suggest an emergentism of the type mentioned above; quite the contrary, [60] takes for granted that the laws of the special sciences *are* deducible from fundamental physics. Rather, the focus of that paper is on the correct definition of terms such as "reducible" and "physical law". According to Fodor, the possibility of multiple deducibility of macroscopic laws via different microphysical mechanisms renders the term "reducible" inappropriate for the special sciences (for example, Newton's law of viscosity, the macroscopic law that describes the resistance of Newtonian fluids to flow, happens to apply to both liquids and gases, although the underlying molecular mechanisms are different for these two classes of fluids); it would be appropriate to speak of "reduction" only if a one-to-one correspondence between the macro- and micro- phenomena was guaranteed. I am not sure I agree with this, but anyway it is mostly a matter of terminology and is not relevant to the topic of the present paper. Nevertheless, what follows from the arguments of the present paper is that while those special sciences that are purely physical (e.g. chemistry, biology, geology, astronomy) are indeed deducible from fundamental physics, those that involve persons (e.g. economics and psychology —the examples mentioned in [60]) are neither completely reducible to (in whatever sense), nor deducible from, fundamental physics, because the phenomena they investigate either include consciousness or are causally affected by free will (Sec. 5). But for an entirely physical system, the physical laws at the microscopic level completely determine the evolution of the positions and velocities (or their probabilities) of the fundamental particles that comprise the system, and therefore completely determine also its macroscopic behaviour; nothing is left out. Hence, strong emergentism would entail that the fundamental physical laws at the microscopic level are violated.

[6] Surprisingly, Davidson, in the same paper[16] that contains the above quotation, claims that there cannot be such laws. However, the definition of supervenience quoted above from his paper guarantees that there are (the only question being whether they are primitive or not), if supervenience physicalism is true. It seems to me that he argued, unconvincingly, against the existence of such laws in an effort to avoid the second hard problem of physicalism which will be discussed in Sec. 3. Quoting again from [16]: "We must conclude, I think, that nomological slack between the mental and the physical is essential as long as we conceive of man as a rational animal". The realisation of this fact could have led him to conclude that physicalism is false, yet he opted to devise an inconsistent version of physicalism instead.

[7] For example, it is my opinion that a large part of the reality we perceive in everyday life is real only in a relative sense. This applies in particular to complex objects. The physical space in which we live is filled with physical particles of matter and energy, the objects of study of fundamental physics, which are not directly observable by us; nevertheless, collections of them have properties that do make them observable. Conceptually, the countless particles existing in our environment could be grouped together in practically infinite ways, but only a limited number of these groups make sense to us, which we perceive as objects. For example, speaking somewhat simplistically for the sake of brevity, a chair, a shoe, a laptop, a car and a house are viewed as objects because they have special meanings for us, due to their utility and their function; a statue and a painting are viewed as objects because their shape or appearance resembles other objects, which is a mind-related role for them; a pebble is viewed as an object because of the uniformity of macroscopic material properties inside it, and the fact that we can handle it as a single object, etc. So, the question of whether Theseus' ship remains the same if we change a plank or a nail does not have an objective answer, because a ship is not a substance but a mind-invented complex object. The answer to the question depends on how we define a ship, which we are free to do however we please. Nevertheless, these complex objects are not entirely fictitious either, because on the one hand they are based on real physical elements and properties (which give them the macroscopic properties that characterise them) and on the other hand they have significance for minds, which are *very* real, affecting their lives in particular real ways. In a way, these complex objects lie in the space between two real extremes, the fundamental physical reality and minds, deriving their relative reality from the absolute realities of these two extremes. A similar view, termed "existential relativity" is expressed in [61], but there it is presented as paradoxical because, as the author is a physicalist, it assumes that persons, relative to whom objects are defined, are themselves also complex objects, thus the definition of objects seems circular. As the reader has probably noticed, my view is that persons are not physical, are not complex, and do not exist "relatively" but absolutely, and are fundamental elements of reality. Hence there is no circularity.

[8] See [[3], chapter 18] for a nice illustration of the basics of the structure and operation of a computer, using mechanical rather than electronic technology.

[9] Theoretically, purpose and teleology, being mental, cannot be included among the fundamental ingredients of a physicalist ontology. But in practice, they are unwittingly implied in most physicalist accounts of the mind. This is most evident in accounts that refer to biological evolution and natural selection (e.g. [29]) which, as if personified, are claimed to have a purpose and act teleologically; thus, ironically, these supposedly Darwinian accounts re-introduce into the ostensibly objective picture of the world what Darwin strived to remove. But also other physicalist theories of mind, such as behaviourism and functionalism, as noted above, rely on concepts such as function, behaviour, information etc. that are ultimately subjective, mind-dependent (even the concept of objects is mind-dependent — see comment [7]). If we invoke the notions of "physical", "design" and "intentional" stances introduced by Dennett [24], the project of physicalism should be to show how the intentional stance is reducible to the physical stance; that this is impossible, the present work aims to demonstrate. Physicalist theories of mind, on the other hand, merely attempt to reduce the intentional stance to the design stance, and fail at that. The design stance, while not using explicitly psychological vocabulary such as "think", "feel", "believe" etc., i.e. vocabulary suggestive of the object studied having itself a first-person perspective, uses vocabulary that

suggests the existence of an outside first-person rational observer that discerns unity (of a composite object), functionality, information, behaviour etc. in that object, things that are subjective and have no place among the foundations of a pure physicalist worldview.

[10] Of course, there are some problems with these assumptions, which are overlooked here because they are not relevant to the point being made. Firstly, in order for there to be understanding, or a feeling of heat or pain, there must be a person who experiences them. Experiences are had by someone, and cannot exist independently of a person. So, the Chinese room, the robot and the car must give rise to *someone*. But who? And why him? What determines this? This sort of questions are very important and will be discussed in Section 4. The other issue is that there does not appear to be a way of knowing with full certainty whether someone other than our own selves exhibits genuine understanding, or feels pain etc., due to the privateness of consciousness. That is, there is no way of really knowing that the Chinese room understands, or that the robot feels pain, or that the car feels heat (unless it happens that we ourselves are the room, or the robot, or the car). But this difficulty can be ignored here.

[11] I cannot exclude the possibility that the intermediate physical causal links that occur inside the brain also cause mental events that constitute components that contribute to our overall unified mental experience of perception.

[12] A physicalist may attempt to circumvent this problem by asserting vaguely that somehow both bodies map to the same person. But this, if it has any meaning at all, would violate physicalism, because the existence of the second body does not in any way physically affect the original body. Therefore, the original body, which produces you, and is unaffected by the construction of the second body, should function exactly the same way as it did before, and hence you, as a product of this unchanged body, should not experience any changes either. In other words, since there is no physical link between the two bodies, there should be no mental link between the persons they produce (and in particular they cannot be the same person).

[13] This thought experiment can also be applied, exactly as it is, against panpsychism (against the claim that each person is the macroscopic aggregate of the alleged micro-consciousnesses of the elemental particles that comprise his/her body).

[14] I apologise to the reader for referring to persons sometimes as male and sometimes as female, due to language restrictions; I do not, in fact, believe persons, at their core, to have a gender. I view the gender as a peripheral quality, or rather a collection of peripheral qualities, which is correlated with biological factors (like other peripheral qualities).

[15] In an effort to avoid the problem, one may contend that in the absence of all peripheral qualities there is nothing left, no person. This argument makes it appear as if a person is just an aggregate of mental phenomena, thoughts, feelings, sensations, etc. By taking a bunch of these and putting them together, one makes a person. This is completely false, as it overlooks the fact that the peripheral qualities do not have an autonomous existence, they cannot be found independently of any existing person, let alone be put together to produce a person. They presuppose the existence of the person who is experiencing them. All mental phenomena are, by definition, experienced by particular minds, and *who* is experiencing them is part of their identity — see also [62].

[16] In other words, let us assume that:

i. I have complete knowledge of how brain structure and processes correlate with mental phenomena. Due to the hard problems of Secs. 2 and 3, the fundamental rules that describe these correlations would have the status of primary, fundamental laws, that are not deducible from the laws of physics, but would themselves be as inexplicable as the fundamental laws of physics themselves.
ii. All mental phenomena have physical correlates in the brain (or body in general).

Under these assumptions (and I am reluctant to accept assumption ii.) I would be able to set all peripheral mental qualities of a person by designing his/her body.

[17] This is the quintessential sense of identity. Some people take identity to be the sum of one's peripheral characteristics so that one's identity changes over time as his/her character changes. But this is a superficial sense of identity, according to which the body-duplicates of the fictitious experiment of section 4.1 would be the same person, when clearly they are not, as anyone can see by imagining that he/she is one of them. Furthermore, even those people who interpret identity in this sense refer to their past and future selves in the first person, not as others; they believe that they deserve praise and admiration for their past virtuous actions and achievements, reward for the work they did in what is now the past, and accept responsibility for their past faults; they plan for their future, etc. Hence even they have a very strong sense of selfhood (which is not what they claim it to be), perhaps subconsciously. Only persons have this sort of identity. Everything else only has a relative identity. In particular, as argued in Comment [7], composite objects exist only relative to mental observers, persons, us, for whom the composite set of constituents has some meaning. It is we that assign an identity to them. As for the fundamental constituents of the material universe (fundamental particles or whatever they may be), unless panpsychists are correct and they have some sort of mental intrinsic nature, they are again without real identity, completely similar in all respects, extrinsic and intrinsic; they are identical from the outside (from the third-person perspective), and there is absolutely nothing in the inside (from the first-person perspective).

[18] As is traditionally customary, I refer to God in masculine gender, but just as for persons I do not regard God as having a gender.

[19] Cf. Matthew 19:23-26, Mark 10:23-27, Luke 18:24-27. In my opinion Jesus in these quotations refers to logical impossibility, as evidenced by his illustrative use of the case of a camel passing through the eye of a needle.

[20] Linda Zagzebski [54][55] has suggested that God is what she calls "omnisubjective", which means that: (a) he has his own, private first-person perspective which is the same as any human first-person perspective, and (b) he has perfect empathy, i.e. he can reproduce (copy) perfectly any person's mental state inside his own consciousness — feel, in his own first-person perspective, what a person feels in their own first-person perspective, and think what a person thinks, imagining himself in that person's place:

> *"If A has perfect total empathy with B, then, whenever B is in a conscious state C, A acquires a state that is a perfectly accurate copy of C and A is aware that her conscious state is a copy of C. A is in this way able to grasp*

> *what it is like for B to be in state C".* [[55], p. 442]

Although this is far better than an omniscience defined as knowledge of all true propositions, in my opinion it still grossly underestimates the nature of God. One aspect of this underestimation is that it purposely maintains the privateness barrier even between God and created persons:

> *"What we do by imagining, God does by directly grasping someone's conscious state ... [O]mnisubjectivity ... does not rely upon the view that two distinct selves literally have the same conscious state. You and God are not confused or merged, and you are not a part of God. If omnisubjectivity is total perfect empathy, it is the most intimate acquaintance possible compatible with a separation of selves".* [[55], p. 443]

If this is so, then the hard problem of creation remains unsolvable even for God, who would find himself in the same embarrassing position as the human creator of Section 4.4 who was endowed with the power of creating persons ex nihilo by a snap of his fingers. He could maybe determine all third-person, descriptive aspects of the created person (intelligence, emotionality, temperament etc.) but not *who* that person would be; for that he would have to rely on factors beyond his control such as, for example, chance. Thus God would not be the source of all reality but merely a super-powerful being who nevertheless is subject to the laws and rules of a greater reality in which he is embedded, and it would be that outside reality that actually determines the identity of the created person, not him. Therefore, in order to do justice to God, I think that we should replace Zagzebski's "He knows what it is like to be you" ([[55], p.443]) with "He knows what it is to be you" — and in fact, He knows it even better than yourself. Furthermore, I disagree with the idea that God experiences life through a first-person perspective that is qualitatively the same as ours. I do not believe that we have the ability to imagine "what it is like" to be God. God is transcendent and his life is mysterious. Omnisubjectivity, on the other hand, is something that we can imagine, wrap our heads around, something that perhaps a human person with psychic abilities could have. It brings God down to to human measures. Finally, I believe that some sort of union with God is the eschatological purpose of our lives, but such a union is precluded if omnisubjectivity is the best that God can do in his relationship with created persons.

[21] The philosophical debate about free will focuses on whether determinism or indeterminism is true. Quantum mechanics tells us that even if physicalism were true, the world would not be entirely deterministic. For the present discussion which focuses on the nature of the mind, epiphenomenalism is a more appropriate notion than determinism, as the question is whether we are governed entirely by physics or not, irrespective of whether that physics is deterministic or includes a degree of randomness.

[22] For agent-causal libertarianism to conflict our scientific theories would require that it offers an alternative explanation to phenomena that are already explained by science. That is, if we had a large body of experimental measurements providing very detailed and accurate pictures of the operation of living human brains at the molecular level (which is a very distant prospect, if ever achievable) and all the data were always in accordance with what the physical laws predict, and yet agent-causal libertarianism claimed that it was not the physical laws but the agent as a substance that causes these

molecular motions, then indeed there would be conflict (although this may still be debatable, as will be shortly mentioned; nevertheless in such a case agent-causality would certainly be almost completely implausible). But that is not what agent-causal libertarianism claims; rather, it claims that the agent/mind has a limited ability to override the physical laws somewhere in the brain by directly causing physical events to happen, when he/she exercises his/her free will. There remains the possibility that the agent has causal power, but he/she always makes (freely) the choice whose footprint on the brain happens to coincide with what the physical laws would determine anyway (if agent-causal power had not been exercised). Perhaps God arranged for this to be the case by setting up the initial conditions of the physical universe in accordance to the free choices that agents would make during their lifetimes, which he knew by his transcendental omniscience even before he created the world. In this case agent-causality would in some sense be true but this would be impossible to test and verify. Although I cannot disprove this view, I find it highly implausible; nevertheless, it seems to be precisely the sort of view held by Immanuel Kant [63].

[23] As a sidenote, it is noteworthy that this statement shows that Spinoza does not consider reasons to be causes; otherwise, everyone would know the cause of their actions, since they know the reasons behind their decisions. He believes that there must be other, hidden, causes in the background, since reasons do not have determining power.

[24] I am limiting the discussion here to physical determinism, but the same argument could be used against any sort of determinism provided that we could, even in principle, know a priori the causes of our actions and their mechanics, be they physical or otherwise, so as to deduce what they determine our actions to be in the future.

[25] I will go into more detail in a separate publication, but a sketch of an explanation for this is the following. For prediction of one physical system by another it is necessary that the predictor have at least as many degrees of freedom (components used for representation) as the predicted system has components; furthermore, these degrees of freedom must evolve at a fast enough pace such that they represent, according to some map, the corresponding degrees of freedom (components) of the predicted system *at a future time*. Self-prediction is then impossible, since the predictor has exactly the same number of degrees of freedom as the predicted system (its own self)— or rather, it has a smaller number, because the representational components must represent their own future selves plus any non-representational, auxiliary components of the system — and while they are required to represent their own selves at a future time, they obviously also represent their current selves, necessitating a match between current and future states, and such a mapping, if required to be fixed, can be established only for trivial cases, such as if the system is idle, doing nothing, so that its future state is the same as the current one.

[26] At this point I anticipate the usual compatibilist (or perhaps even hard incompatibilist) reaction that this reasoning conflates determinism with fatalism (e.g. [64][65]). But it should be obvious that it is whoever raises this objection that conflates fatalism with determinism. It is obvious that in the above deterministic scenarios, since the specified events did occur (whether mental, like thinking about and deciding to become evil, or physical, like the actual action of harming someone) they were determined by the past and the laws. If determinism is true, nothing can occur unless it was determined to occur by the past and the laws. Fatalism, on the other hand, is meaningful only in indeterministic universes, or rather in universes where minds have free will, and where various paths of action are available to agents to freely

choose from, and yet they all lead to the same outcome. Determinism could be viewed as a limiting case of all-out fatalism, where it is not only certain events that are fixed but everything, including the will of the agents. But this could cause misunderstandings, as fatalism is usually perceived as something happening in spite of something else, whereas in determinism there is no "in spite of" something, but everything proceeds according to the laws. It seems to me that there is a lot of confusion as to what determinism really means and what it entails, even among philosophers who preoccupy themselves with this issue. For example, in [65] fatalism is defined as the idea that "outcomes are not dependent on agents' actions or intentions" (p. 91-92). In determinism, outcomes *are* dependent on agents' actions and intentions, but these actions and intentions are themselves determined by the past and the laws, by factors outside of the control of the agent. The agent is not the ultimate source of what he/she will think, will, intend, decide, and do. In a sense, of course, what they will do depends on what they will decide, and what they will decide depends on what they will think, but all of these are determined by the state of the universe in the distant past, when the agent did not even exist, and laws such as Newton's laws of motion, Maxwell's laws of electromagnetism etc. In exploring different degrees of fatalism, Holton, a believer in determinism who tries to clarify its distinction from fatalism, writes:

> *"[Fatalism means that t]here are some outcomes such that whatever action Oedipus were to perform, they would come about. ... We could go further and imagine a global action fatalism, where Oedipus's choices would have no impact on his actions, even the most basic. In other words, whatever Oedipus chose, his actions would be the same. ... Could we go further still, and imagine a form of fatalism in which Oedipus had no control even over his choices? Here things become unclear. We have made sense of Oedipus's powerlessness in terms of the lack of impact of something that is in his control on something else: his choices and actions lack the power to affect certain outcomes. Once he loses control of everything, we can't see it this way; and once he loses control of his choices, it looks as though he has lost control of everything. Perhaps there is some way of understanding a totally global fatalism, but I don't see what it is." ([65], p. 88).*

So here we have an admission by Holton that he does not understand determinism, in his own words, since determinism is precisely what he describes as "totally global fatalism". This lack of understanding of determinism is most likely much more widespread than one would expect — how else can one make sense of the widespread adoption of the absurd doctrine of compatibilism?

[27] It is noteworthy that most hard incompatibilists accept these notions as objective aspects of reality, and find them incompatible with determinism. That is, they understand what moral responsibility (and the rest of these notions) is, they agree that if we had libertarian free will then we would have moral responsibility, but they believe that we do not have libertarian free will because it is an empirical fact that reality is deterministic. But then the question arises: if reality is deterministic and free will is an illusion, why do these notions exist objectively in the first place? And how are we humans — deterministic creatures that lack free will — able to understand them?

[28] If it is not deliberately that they are forgetting it then this means that they do not fully understand what determinism is all about. This is actually something very likely, as the details and the arguments of the compatibilist-incompatibilist debate

reveal much confusion. For example, it is very surprising to me that the "consequence argument", which merely states the obvious, has had a large impact in the debate [[46], Chapter 4]. This cannot be explained unless determinism was (and probably still is) misunderstood.

[29] This is similar to a physicalist intuitive fallacy whereby since (1) physicalism is true, and (2)we have consciousness, it follows that consciousness is physical.

[30] For those who are having difficulty understanding the precise meaning of epiphenomenalism, maybe this example will help. Suppose a scenario where physicalism is true. We set up a sequence of standing dominoes each of which contains an artificial, electronic brain that generates consciousness. Using optical sensors, each domino is consciously aware of the domino that is placed directly in front of it. We initiate a domino run by toppling the first domino. Each domino falls on the next one and topples it, in a deterministic process governed by gravity, Newton's laws of motion etc. But furthermore, we have programmed the dominoes' chips such that when impacted by another domino they consciously feel an urge to fall onto the domino in front of them. This is achieved by placing transducers on the dominoes' surface which detect the pressure from impact and send electrical signals to the brain chip which activate the urge-feeling circuitry. In addition, we have equipped each domino with a small cavity partially filled with fluid which, with the aid of appropriate sensors, measures the tilt angle of the domino (it is a system similar to the vestibular apparatus in our ears). When the tilt angle increases from zero to a small value of, say, five degrees, the system is designed such that a signal is sent to the chip that activates a piece of circuitry that generates a conscious feeling that the domino has decided to fall onto the next domino. Thus, when its time to fall comes, the domino suddenly feels an urge to fall onto the domino in front of it, and almost immediately (has the illusion that it) decides to do so. The falling ensues. Suppose finally that we have programmed the dominoes so that they prefer the upright position to the one lying down, perhaps using the same fluid-filled cavity to send a signal to the brain to generate an unpleasant feeling at high tilt angles. According to compatibilism, each domino is culpable for making its neighbour feel unpleasant. However, an understanding of the mechanics behind the behaviour of the dominoes makes this claim absurd, at least if right and wrong are assumed to be objective aspects of reality. However, compatibilists could be justified in their claim if right and wrong are not objective aspects of reality but are mere human inventions, conventions, or subjective feelings. In this case we would be free to *decide* that the dominoes are culpable, based on a suitably chosen definition of culpability (e.g. someone is culpable for some event if all of the following hold: (a) he/she felt an urge for causing it, (b)he/she felt that he/she decided to cause it, (c) the event occurred and (d) he/she was physically involved in the event). If culpability is a man-made concept then we are free to define it as we please.

[31] Of course, we observe other persons with our senses, but only indirectly, through their physical bodies, whereas they themselves are directly inaccessible to us, and we have to make inferences about them and their mental states based on the direct experience we have of our own selves (with some help from hereditary instinct which informs us about their mental state from their physical appearance — e.g. even babies seem to recognise a human face and its expressions).

[32] The discrepancy between physics and logic on the one hand and ethics on the other hand is actually somewhat less sharp than portrayed here, because even physics and logic, despite referring to the outside world, are ultimately

understood by minds, privately, and so there can be a degree of subjectivity and uncertainty, as Descartes notoriously noted. Nevertheless it is undeniable that the input from our senses about the physical world makes it much more compelling to accept physical and logical truths than what is possible for ethical theses. As I said in Sec. 5, I do not believe that moral "laws" are fundamental but they derive from the infinite intrinsic value of persons/minds (Matthew 22:36-40). Now, the moral "laws" are built on top of that, being particularisations, for specific circumstances of life, of the ethical principle of valuing and loving persons, and they do have a component that is based on reason and logic, which is the part that connects this particular set of circumstances with the consequences that they have on persons — e.g. murder will terminate one's biological life; theft will deprive them of possessions that they worked hard and spent time to acquire and which they need; physical abuse will cause pain and psychological trauma etc. This part of the moral "laws" depends on the circumstances of life and on the way life is perceived by society and can be different from one culture to another; in a different universe (e.g. where persons are immortal, do not have physical possessions etc.), the moral laws could be entirely different than our own, although their foundation would be the same. This foundational core principle of valuing someone, loving them, having their best interest in mind, is not provable by logic, any more than is the opposite principle of prioritising self-interest. We have an intuitive comprehension of what it is for someone to be harmed or benefited, stemming from our own introspection and experience of our own selves, and we are free to chose our moral path. It is for this reason that our free will's main manifestation is in relation to ethics. Moral nihilism is not logically disprovable, so that one can be justified in making both moral and immoral decisions and actions, as argued in Section 5. Reasons, therefore, in general, do not determine whether we will make a moral or immoral choice, but we choose freely which reasons we want to value most, so that our freedom is mostly manifested in the moral realm.

[33] Of course, then the problem faced by Sosa[61] arises, namely that persons exist subjectively as perceptions of persons, and therefore their existence has a circular dependency on its own self.

[34] Cf. Matthew 22:39, Mark 12:31.

[35] Provided that back then I (as a mental substance) was already joined with the physical mass that constituted the fetus that was to evolve into my body. Since personhood is not derivable from the physics of the body, there is nothing that necessitates that the union of the person with his/her body occurs at biological conception and not later. In fact there likely is no way of knowing when the union occurs, due to privateness (the "problem of other minds").

[36] In particular, I find modal arguments employing the dubious concept of "possible worlds" as of very limited value.

[37] Cf. John 3:12, 1 Corinthians 2:9, 1 John 3:2.

References

1. ^*Berryman, Sylvia. Democritus. Edited by Edward N Zalta, Winter 2016 ed., Metaphysics Research Lab, Stanford University, 2016.*
2. ^*Dennett, Daniel C. Consciousness Explained. Little, Brown and Company, 1991.*

3. [a, b]*Stewart, Ian. Concepts of Modern Mathematics. Dover Publications, 1995.*
4. [a, b]*Chalmers, David J. Consciousness and Its Place in Nature. Edited by Stephen P Stich and Ted A Warfield, John Wiley & Sons, Ltd, 2003, pp. 102–142.*
5. [a, b]*Goetz, S. (2001). Modal Dualism: A Critique. In K. J. Corcoran (Ed.), Soul, Body, and Survival: Essays on the Metaphysics of Human Persons (pp. 89–104). Cornell University Press.*
6. [^]*Taliaferro, C., & Evans, J. (2011). The Image in Mind: Theism, Naturalism, and the Imagination. Bloomsbury Publishing.*
7. [a, b]*Feser, E. (2006). Philosophy of mind: A beginner's guide. Oneworld Publications.*
8. [a, b]*Chalmers, D. J. (1995). Facing up to the problem of consciousness. Journal of Consciousness Studies, 2(3), 200–219.*
9. [^]*Descartes, R. (1988). Descartes: Selected philosophical writings. Cambridge University Press.*
10. [^]*Kirk, R., & Squires, R. (1974). Zombies v. materialists. Proceedings of the Aristotelian Society, Supplementary Volumes, 48, 135–163.*
11. [a, b]*Jackson, F. (1982). Epiphenomenal qualia. The Philosophical Quarterly, 32(127), 127–136.*
12. [a, b, c]*Nagel, Thomas. "What Is It like to Be a Bat?" The Philosophical Review, vol. 83, 1974, pp. 435–450.*
13. [^]*Levine, Joseph. "Materialism and Qualia: The Explanatory Gap." Pacific Philosophical Quarterly, vol. 64, 1983, pp. 354–361.*
14. [^]*Casalino, L., Gaieb, Z., Goldsmith, J. A., Hjorth, C. K., Dommer, A. C., Harbison, A. M., … Amaro, R. E. (2020). Beyond Shielding: The Roles of Glycans in the SARS-CoV-2 Spike Protein. ACS Central Science, 6(10), 1722–1734.*
15. [a, b]*Horgan, Terence. "From Supervenience to Superdupervenience: Meeting the Demands of a Material World." Mind, vol. 102, 1993, pp. 555–586.*
16. [a, b, c]*Davidson, Donald. (1970). Mental events. Experience and Theory, vol. 79--101 . Foster, L. and Swanson, J. W. University of Massachusetts Press..*
17. [a, b]*Goff, P., Seager, W., & Allen-Hermanson, S. (2021). Panpsychism. Retrieved from The Stanford Encyclopedia of Philosophy website: https://plato.stanford.edu/archives/win2021/entries/panpsychism/*
18. [^]*Chalmers, D. J. (2015). Panpsychism and panprotopsychism. In T. Alter & Y. Nagasawa (Eds.), Consciousness in the physical world: Perspectives on Russellian monism (pp. 246–276). Oxford University Press Oxford.*
19. [^]*Mørch, H. H. (2017). Is Matter Conscious? Nautilus, 47, 90–96.*
20. [a, b]*Chalmers, D. J. (2017). The combination problem for panpsychism. In G. Brüntrup & L. Jaskolla (Eds.), Panpsychism: contemporary perspectives (pp. 179–214). Oxford University Press.*
21. [^]*Brentano, F. ([1874,] 2012). Psychology from an empirical standpoint. Routledge.*
22. [^]*Chisholm, R. M. (1957). Perceiving: A philosophical study.*
23. [a, b]*Dretske, F. (2002). A Recipe for Thought. In D. J. Chalmers (Ed.), Philosophy of Mind: Classical and Contemporary Readings (pp. 491–499). Oxford University Press.*
24. [a, b, c, d, e, f]*Dennett, D. C. (1981). True believers: The intentional strategy and why it works. In A. F. Heath (Ed.), Scientific Explanation: Papers Based on Herbert Spencer Lectures Given in the University of Oxford (pp. 150–167). University of Massachusetts Press.*

25. [a, b]*Adams, F., & Aizawa, K. (2021). Causal Theories of Mental Content. Retrieved from The Stanford Encyclopedia of Philosophy website: https://plato.stanford.edu/archives/fall2021/entries/content-causal/*
26. [a, b]*Horgan, T., & Tienson, J. (2002). The intentionality of phenomenology and the phenomenology of intentionality. In D. J. Chalmers (Ed.), Philosophy of mind: Classical and contemporary readings (pp. 520–533). Oxford University Press.*
27. [^]*Cole, D. (2020). The Chinese Room Argument. Retrieved from The Stanford Encyclopedia of Philosophy website: https://plato.stanford.edu/archives/win2020/entries/chinese-room/*
28. [a, b, c]*Reppert, V. (2009). The argument from reason. In W. L. Craig & J. P. Moreland (Eds.), The Blackwell companion to natural theology (pp. 344–390). John Wiley & Sons.*
29. [a, b]*Millikan, R. (1989). Biosemantics. Journal of Philosophy, 86, 281–297.*
30. [^]*Dretske, F. (1988). Explaining behavior: Reasons in a world of causes. MIT press.*
31. [a, b, c]*Artiga, M., & Sebastián, M. Á. (2020). Informational theories of content and mental representation. Review of Philosophy and Psychology, 11(3), 613–627.*
32. [^]*Kim, J. (2005). Physicalism, or something near enough. Princeton University Press.*
33. [^]*Foster, J. (1991). The immaterial self: A defence of the Cartesian dualist conception of the mind. Routledge.*
34. [^]*Audi, P. (2011). Primitive causal relations and the pairing problem. Ratio, 24(1), 1–16.*
35. [a, b]*James, W. (1890). The Principles of Psychology. New York: Henry Holt.*
36. [^]*Long, H. S. (1948). Plato's Doctrine of Metempsychosis and its Source. The Classical Weekly, 41(10), 149–155.*
37. [a, b, c, d]*Moreland, J. P., & Craig, W. L. (2017). Philosophical foundations for a Christian worldview, 2nd edition. InterVarsity Press.*
38. [^]*Erasmus, J. (2018). The Kalām Cosmological Argument: A Reassessment. Springer.*
39. [a, b]*Erasmus, J., & Verhoef, A. H. (2015). The Kalām cosmological argument and the infinite God objection. Sophia, 54, 411–427.*
40. [^]*Wierenga, E. (2021). Omniscience. Retrieved from The Stanford Encyclopedia of Philosophy website: https://plato.stanford.edu/archives/sum2021/entries/omniscience/*
41. [^]*Kane, R. (2016). On the role of indeterminism in libertarian free will. Philosophical Explorations, 19(1), 2–16.*
42. [^]*De Caro, M., & Putnam, H. (2020). Free Will and Quantum Mechanics. The Monist, 103(4), 415–426.*
43. [^]*Chisholm, R. (1964). Human Freedom and the Self. The Lindley Lectures.*
44. [^]*Davidson, D. (1963). Actions, Reasons, and Causes. Journal of Philosophy, 60(23), 685.*
45. [a, b]*Plantinga, A. (1993). Warrant and proper function. Oxford University Press.*
46. [a, b, c, d, e]*McKenna, M., & Pereboom, D. (2016). Free will: A contemporary introduction. Routledge.*
47. [a, b, c]*Pereboom, D. (2014). Free will, agency, and meaning in life. Oxford University Press.*
48. [a, b]*Pereboom, D. (2001). Living Without Free Will. Cambridge University Press.*
49. [^]*Silverthorne, M., & Kisner, M. J. (2018). Spinoza: Ethics: Proved in Geometrical Order. Cambridge University Press.*
50. [^]*Rummens, S., & Cuypers, S. E. (2010). Determinism and the Paradox of Predictability. Erkenntnis, 72(2), 233–249.*
51. [^]*McKenna, M. (2008). A hard-line reply to Pereboom's four-case manipulation argument. Philosophy and Phenomenological Research, 77(1), 142–159.*

52. ^*Dennett, D. (2015). Stop telling people they don't have Free will. Big Think. Retrieved from https://bigthink.com/videos/daniel-dennett-on-the-nefarious-neurosurgeon/*
53. ^*Street, S. (2006). A Darwinian dilemma for realist theories of value. Philosophical Studies, 109–166.*
54. a, b*Zagzebski, L. (2008). Omnisubjectivity. In J. Kvanvig (Ed.), Oxford Studies in Philosophy of Religion (pp. 231–248). Oxford University Press.*
55. a, b, c, d, e*Zagzebski, L. (2016). Omnisubjectivity: why it is a divine attribute. Nova et Vetera, 14(2), 435–450.*
56. ^*Hasker, William. Persons as Emergent Substances. Edited by Kevin J Corcoran, Ithaca: Cornell University Press, 2001.*
57. ^*Kim, J. (2001). Lonely souls: Causality and substance dualism. In K. J. Corcoran (Ed.), Soul, Body, and Survival: Essays on the Metaphysics of Human Persons. Ithaca: Cornell University Press.*
58. ^*Leftow, B. (2001). Souls Dipped in Dust. In K. J. Corcoran (Ed.), Soul, Body, and Survival: Essays on the Metaphysics of Human Persons (pp. 120–138). Cornell University Press.*
59. a, b*Broad, C. D. (1925). The Mind and its Place in Nature. Routledge.*
60. a, b, c*Fodor, J. (1974). Special Sciences (Or: The Disunity of Science as a Working Hypothesis). Synthese, 28, 97–115.*
61. a, b*Sosa, E. (1999). Existential relativity. Midwest Studies in Philosophy, 23, 132–143.*
62. ^*Lowe, E. J. (2014). Why my body is not me: The unity argument for emergentist self-body dualism. In A. Lavazza & H. Robinson (Eds.), Contemporary Dualism: A Defense (pp. 245–265). Rootledge.*
63. ^*Pereboom, D. (2006). Kant on Transcendental Freedom. Philosophy and Phenomenological Research, 73(3), 537–567.*
64. ^*Holton, R. (2009). Determinism, self-efficacy, and the phenomenology of free will. Inquiry, 52(4), 412–428.*
65. a, b, c*Holton, R. (2013). From determinism to resignation; and how to stop it. In A. Clark, J. Kiverstein, & T. Vierkant (Eds.), Decomposing the Will (pp. 87–100). Oxford University Press.*