International Studies in the Philosophy of Science

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/cisp20

Hempel's Dilemma: Not Only for Physicalism*

Erez Firt, Meir Hemmo & Orly Shenker

INTERNATIONAL STUDIES IN THE PHILOSOPHY OF SCIENCE
https://doi.org/10.1080/02698595.2022.2041969

Routledge
Taylor & Francis Group

Check for updates

RESEARCH ARTICLE

Hempel's Dilemma: Not Only for Physicalism*

Erez Firt[a], Meir Hemmo[a] and Orly Shenker[b]

[a]The University of Haifa; [b]The Hebrew University of Jerusalem

ABSTRACT
According to the so-called Hempel's Dilemma, the thesis of physicalism is either false or empty. Our intention in this paper is not to propose a solution to the Dilemma, but rather to argue as follows: to the extent that Hempel's Dilemma applies to physicalism it *equally* applies to *any* theory that gives a *deep-structure* and *changeable* account of our experience or of high-level theories. In particular, we will show that it also applies to mind–body dualistic theories. The scope of Hempel's Dilemma turns out to be very wide: it is a special case of a general sceptical argument against changeable deep-structure theories in and outside science.

KEYWORDS
Hempel's Dilemma; deep structure; dualism; physicalism; reduction; theory change

1. Introduction

> Physicalism is the thesis that everything is physical, or as contemporary philosophers sometimes put it, that everything supervenes on the physical. … Of course, physicalists don't deny that the world might contain many items that at first glance don't seem physical — items of a biological, or psychological, or moral, or social nature. But they insist nevertheless that at the end of the day such items are either physical or supervene on the physical. (Stoljar 2017)

But what is it for something to be 'physical'? On the one hand, physicalism is a metaphysical doctrine, in that it is about what there is in the world. On the other hand, it has a close relation to the physical sciences, expressed by the claim that what exists in the world is what physics maintains exists.[1] This theory-dependence of the physicalist metaphysical approach troubled Carl Hempel (1980), who in response formulated a claim regarding the scope of the language of physics. Hempel wrote:

> I would add that the physicalistic claim that the language of physics can serve as a unitary language of science is inherently obscure: The language of what physics is meant? Surely not that of, say, 18th century physics; for it contains terms like 'caloric fluid', whose use is governed by theoretical assumptions now thought false. Nor can the language of contemporary physics claim the role of unitary language, since it will no doubt undergo further changes, too. (Hempel 1980, 194–195)[2]

CONTACT Orly Shenker orly.shenker@mail.huji.ac.il
*This paper is dedicated to the memory of Professor Hugh Mellor, who co-authored the paper (Crane and Mellor 1990) that brought Hempel's Dilemma to the attention of philosophers interested in reductive projects in general and physicalism, in particular, and thus inspired the present paper.

And what about *future* physics? Hempel, in the above quotation, does not address this option. It became, however, part of what is now known as 'Hempel's Dilemma', following the work of Crane and Mellor. In their 1990 paper titled 'There is no Question of Physicalism', Crane and Mellow referred to the reduction of different scientific fields to physics and famously described the problem that the Dilemma poses as follows:

> We must first ask to what physics the RIP [Reduction in Principle] principle is supposed to be applied: to present physics, or to some hypothetical future physics? This question poses a dilemma. For applying the principle to present physics entails that any future extensions of it would not be physical: that physics, the paradigm physical science, is already complete. But no one believes this. And if we apply the principle to an otherwise unspecified future physics, we shall not be able to say which sciences are physical until we know which of them that physics must cover - which is just what the principle was supposed to tell us. To use RIP to future physics to say what that physics must cover if it is to cover everything physical is obviously viciously circular. So the physical cannot be defined as what is reducible in principle to physics, either present or future. (Crane and Mellor 1990, 188)[3]

Since then the Dilemma has been discussed extensively and described in various ways.[4] The differences between the formulations, their merits and demerits, are *not* the focus of our argument here (although they are of course important for other matters), and therefore for our purpose it is enough to have in mind any of the main proposed formulations (e.g. Crane and Mellor 1990; Crook and Gillett 2001; Dowell 2006b; Gillett and Witmer 2001; Hellman 1985; Montero 1999; Stoljar 2010; Wilson 2006; we give a brief survey of the literature in Section 2). Here is a rough convenient formulation for capturing the main points important for our purposes; it is not identical to any of the formulations by any of the writers that addressed this Dilemma, but presents the main gist common to some of them:

On the one hand, there are good reasons to think that contemporary physics (let alone past physics) is false, at least to some extent; on the other hand, the nature of future physics is, to a significant extent, unknown. Therefore the *reference* (or more generally *meaning*) of 'physics' is either *false* or so ambiguous as to be, for all practical purposes, *empty*. Therefore physicalism is unacceptable.

The two horns of the Dilemma are closely related. The main reason behind both is this. Problems within contemporary physics lead many to suspect that the current theories of physics will change or be replaced by others. Some thinkers conjecture that these changes may be radical in a way that is at present unforeseeable (second horn of the Dilemma). Others think that although current physics may undergo radical changes in some domains in the future (e.g. high energy physics, quantum gravity), these changes will most likely be irrelevant to accounting for the mind (e.g. Smart 1978; Bokulich 2011). Still others think that, regardless of the mind, current physics is on the right track and unlikely to change radically. We tend to agree with the first view and disagree with the two latter ones (see some examples and discussion in Section 3.2).

Of course, in addition to the need to solve the known open problems in contemporary physics, one may employ the 'pessimistic induction' line of thinking according to which the fact that physics underwent considerable changes in the past gives reason to believe, or at least to remain open to the possibility, that it will undergo considerable changes in the future.

In this paper our aim is not to propose a solution to Hemple's Dilemma, but rather to argue as follows: to the extent that Hempel's Dilemma applies to physicalism it *equally* applies to *any* theory that attempts to give a *deep-structure* and *changeable* account of our experience. In particular, we will argue that Hemple's Dilemma applies not only to physicalism, but and to the same extent to mind–body dualistic theories, provided the latter attempt to give a deep-structure account of our experience. Our conclusion is that the scope of Hempel's Dilemma turns out to be much *wider* than usually thought: the Dilemma is a special case of a general sceptical argument against deep structure and changeable theories in and outside science.

The paper is structured as follows. In Section 2, we briefly set the context and describe the main proposals in the literature addressing Hempel's original Dilemma. Later on we will see that none of them successfully resolves the generalised version of the Dilemma. In Section 3, we present two conditions (or requirements) for the applicability of Hempel's Dilemma: one (Section 3.1, 3.3) is that the theories it addresses need to be deep structure theories; and the other (Section 3.2) is that the theories must be subject to possible (relevantly significant) change. Taking these two requirements as by and large sufficient for the applicability of the Dilemma, we come up with a generalised Hempelian Dilemma (in Section 3.4). In Section 4, we show that the generalised Hempelian Dilemma applies not only to physicalism, but also to outright psychophysical dualism (Section 4.1). We also show that the generalised Hempleian Dilemma applies in the same way to non-reductive physicalism (Section 4.2), and also to non-fundamental theories (Section 4.3), that is, for example, biology. A consequence is that our generalised Dilemma applies also to non-foundationalist metaphysics. Finally in Section 5, we explore the way in which identity statements are related to the generalised Dilemma: we show which of them are subject to Hempel's Dilemma and which are not. We conclude (Section 6) by pointing out that, following our analysis, Hempel's Dilemma ceases to be a threat specifically to physicalism; instead, it turns out to be part of a general sceptical argument against changeable deep structure theories in and outside science.

2. A Brief Survey of Some Proposed Solutions to Hempel's Dilemma

Poland (2003, 30) expresses the worry that many writers share concerning the significance of the Dilemma and the need to solve it:

> It is of some importance, then, that although physicalism apparently has profound significance and has been in the spotlight to a considerable extent, proponents of physicalism have not adequately responded to a frequently voiced objection, viz., that the theses of physicalism are vacuous because the critical concept of the physical has no well-defined content. Although stated in different ways by different critics (cf. Feigl 1969, 21; Hempel 1980, 194–5; Crane and Mellor 1990, 186–7), the core objection is a deep and rather complicated one with far-ranging implications for physicalism. If the theses of physicalism are vacuous, then they cannot properly be conceived of as empirical hypotheses which have a truth value and which play a role in science and philosophy. And if the theses are vacuous, it is unclear how physicalism can have any human significance.

Supporters of physicalism proposed various ways to either overcome the original Dilemma or live with it. Let us describe in outline and very briefly the main proposals or directions of responses to the Dilemma in the literature. We shall not go into details or critical discussions of them here. Our main point is this: since (arguably; see below) there is a generalised Hempelian Dilemma that is applicable to any changeable deep-structure account of our experience, not only to an account by physics, the question arises whether any of the proposals that attempt to resolve the original Dilemma for physicalism may also successfully resolve the generalised one.

For convenience the main proposals in the literature can be divided into the following major groups.

One group of proposals claims that it is reasonable to think that the ontology is well described by contemporary physics, at least approximately (after all, the success of contemporary physics is a salient motivation for accepting physicalism in the first place). According to these views, physics will not change much, so they reject the claim that we don't know the nature of future physics, thus rejecting the claim of Horn 2. Some in this group also reject the claim of Horn 1, believing that current physics does explain the phenomena satisfactorily, and to the extent that future physics will significantly change, these changes will not affect the part of physics that is relevant to those explanations (see e.g. Smart 1978; Lewis 1994; Bokulich 2011). By contrast, Melnyk (1997, 2003) who also rejects Horn 1 of the Dilemma makes no claim that current physics is even approximately true; instead, he argues that current physics is more likely to be true relative to its rivals, where an as yet unformulated future theory doesn't count as a relevant rival.

A *second* group rejects Horn 2 of the Dilemma and argues that referring to some future physics, that is yet unknown, is the right move, and it is neither trivial nor empty or vague. Among the adherents of this view, Poland (1994), for example, conjectures that the integration of the mental into its future-physical place will be carried out in a similar manner to the incorporation of the electromagnetic theory into fundamental physics, and therefore although we don't know anything about future physics, we do know that current physics will find the place in which the mental can be integrated into it in this sense.

A *third* proposal to characterise 'physics', called *via negativa*, was put forward in the context of psycho-physical reductionism, and characterises the 'physical' as non-mental, thus avoiding the need to address the details of the science of physics, either present or future (see Montero 1999). These views are based on the idea that physics is causally complete (or closed), roughly that physical events are preceded by physical causes (see Spurrett and Papineau 1999; criticism by Gillett and Witmer 2001, and a rebuttal by Montero and Papineau 2005).

In the above quotations from Hempel and from Crane and Mellor, physicalism is addressed without saying what kind of facts are to be reduced to physics. Yet, Hempel's Dilemma is often discussed in the specific context of mind–body physicalism, and it is here that the via negativa comes into play. Of course, also when examining the reduction to physics of biological phenomena or high-level theoretical facts (entities, properties, etc., from other special sciences), one can employ the via negativa, saying that 'physical' is everything that is not 'biological', but that may sound strange. Why is that so? In the context of the mind–body problem the via negativa seems natural because the intuitions concerning the mental are still different from those concerning the biological. To see this, consider the views in previous times, when the option that life involves some sort of *elan vital* was still on the table. In those times it would have been perhaps more natural to characterise the physical as whatever is not life (or living, or whatever the case may be). Montero (1999) replaces the distinction mental-phyterosical with the distinction mental-nonmental, and so she does not seem to identify the physical with nonmental But Worley (2006) defends the via negativa approach via Montero 's revised distinction arguing that our concept of the physical does includes a built-in contrast with the mental as one of its components. We briefly discuss Monetro's view in Section 4.1.

A *fourth* approach attempts to find definite characteristics of physics, such as the fact that physics describes the most fundamental elements of the universe, and perhaps also the fact that physics is the only kind of scientific theory in which the laws are strict and have no ill-defined exceptions in the sense that also includes genuine probabilistic laws (see Dowell 2006a for an approach in this direction). However, it is unclear that such an approach can avoid the addition of something along the lines of the via negativa approach in order to block the possibility that facts about the mental are incorporated into physics as primitives (see Wilson 2006).

A *fifth* group tries to avoid the Dilemma altogether by denying that physicalism is a sentence with a truth value (or an expression of a belief about the world that may be true or false); thus in a sense embracing the Hempelian conclusion and its close relative, the so-called Chomsky's challenge (see Chomsky[5] 1968, 2000, 2003; Poland 2003) that physicalism is vacuous because the concept of the physical lacks content. Instead, this approach advances the idea that physicalism means adopting an *attitude* or a stance to form one's theory of the world according to what the best available theories of physics at the time say exists. Supporters of this view in different versions are Hellman (1985); van Fraassen (2002)[6]; Poland (2003); and Ney (2008b). In his criticism of physicalism and response to Poland (2003), Chomsky (2003, 265) seems to support this view: 'One can entertain the idea that "the mental is the neurophysiological at a higher level," but for the present, only as a guide to inquiry, without much confidence about what "the neurophysiological" will prove to be'. Here the question arises of what sort of arguments may justify adopting this stance rationally and non dogmatically; see Stoljar (2017, Sec. 12) for additional problems.

In this paper, we propose to look at the Dilemma from yet another perspective. Our proposal is this: *Hempel's Dilemma is not solved nor solvable; and nevertheless it is not a threat to physicalism*. The reason for this is that the domain of application of the generalised Dilemma is much wider: it applies to all the theories that explain the phenomena or high-level facts by appealing to some underlying deep structure, and that use methodologies that allow for a change of these explanations. Physicalism indeed offers a changeable deep structure account of the high-level facts, and is therefore subject to Hempel's Dilemma. But in our view, as we will argue, mind–body dualism also offers changeable deep structure accounts and is therefore no less subject to Hempel's Dilemma. Notice that this is a *novel* charge which arises in addition to the usual charge against dualism in the context of Hempel's Dilemma according to which dualism includes the claim that some things are physical and fundamentally so (and some aren't), but then the Dilemma arises with respect to what 'physical' means in more or less the same way that it arises for physicalism.

An anonymous reviewer of this paper summarised our claim as follows:

> Suppose we make a metaphysical claim (physicalism, dualism, whatever) and the claim uses a category 'X'. Either 'X' is understood by reference to some theory that is changeable and hostage to empirical fortune or it's not. If it is, then a dilemma arises for it, since any current theory is surely not to be trusted, and any ideal theory is surely unknown. The only alternative is to appeal to an understanding of "X" that doesn't depend on such an empirical theory.

So the claim is indeed not complicated to grasp. For one reason or another, however, this is not the way that Hempel's Dilemma is seen in the literature: it is taken to be a threat *specifically* for physicalism, and is therefore used in comparing between physicalism and its rivals. Our aim is to place Hempel's Dilemma in a different context and show that it is not a threat to physicalism as opposed to its rivals, but rather the Dilemma is a way of expressing *general sceptical worries*. Needless to say, such worries are deep and extremely significant for the way we understand science (as well as any other account of the phenomena). But this means that Hempel's Dilemma should be addressed as such, which is quite different from the way it is discussed in the literature.

3. Analysis of Hempel's Dilemma: Requirements for Its Applicability

We now turn to arguing that Hemple's Dilemma applies to any theory that satisfies two quite general requirements that, as we shall see, are not specific to physicalism. This will form the basis for our generalisation of the Dilemma.

3.1. Requirement 1: Deep Structure Theories

The physicalist account of the relation between mind and body is that the mental phenomena and related high-level theoretical facts (including entities, processes, regularities, properties, etc.) are accounted for by the facts (*ditto*) of the low-level theories of physics. We propose to generalise the Dilemma for all the accounts of phenomena or of theoretical high-level facts (*ditto*), by theoretical lower-level (purported) facts (*ditto*).[7] This topic is standardly addressed under headings such as 'metaphysical reduction' and 'theory reduction' (see Shapiro 2018; van Riel and Van Gulick 2019). There are various approaches concerning the relationship between the physical facts and those that are to be reduced to them, and we stress that our argument holds for all of them: we propose that *for any account of the relationship between physics and the facts to be reduced to it, there is a corresponding Hempelian Dilemma.*

Still, in order to be more precise and less abstract, and for the purpose of our argument here, in the context of Hempel's Dilemma, we find it *useful* to focus on one approach in this context, and we choose to employ a distinction offered by Einstein (1919) which is 'Einstein's most original contribution to twentieth-century philosophy of science' (Howard and Giovanelli 2019, §6; for the role of this distinction in Einstein's thinking about the 'logic of discovery', see Giovanelli 2020).[8] Einstein tried to explain what kind of contribution to science his Special Theory of Relativity makes to our knowledge of the world. To do so he compared the status of this theory to the status of other theories of science, in particular thermodynamics and statistical mechanics (to which he also made crucial contributions, e.g. in his 1905b). Reflecting on the relations between different theories in physics, Einstein proposed to distinguish between what he calls '*Principle theories*' and '*Constructive theories*', and offered a non-trivial classification of

Special Relativity as a case of the former rather than the latter, a classification that is non-standard even more than a century later.[9] Einstein explained this distinction in a short essay in the 1919 London Times, as follows:

> There are several kinds of theory in Physics. Most of them are constructive. These attempt to build a picture of complex phenomena out of some relatively simple proposition. The kinetic theory of gases, for instance, attempts to refer to molecular movement the mechanical, thermal and diffusional properties of gases. When we say that we understand a group of natural phenomena, we mean that we have found a constructive theory which embraces them. But in addition to this most weighty group of theories there is another group consisting of what I call theories of principle. These employ the analytic, not the synthetic method. Their starting-point and foundation are not hypothetical constituents, but empirically observed general properties of phenomena, principles from which mathematical formulae are deduced of such a kind that they apply to every case which presents itself. Thermodynamics, for instance, starting from the fact that perpetual motion never occurs in ordinary experience, attempts to deduce from this, by analytic processes, a theory which will apply in every case. The merit of constructive theories is their comprehensiveness, adaptability, and clarity, that of the theories of principle, their logical perfection, and the security of their foundation.

'*Principle theories*', according to Einstein, start off from empirically observed 'general properties of phenomena' and 'deduce' from them mathematical formulae. This characterisation is, of course, extremely imprecise, to such an extent that it is patently flawed: only specific facts can be observed, and the problem of induction prohibits deduction of mathematical formulae from them. An example that may be central to Einstein's (1919) analysis, as well as to our present discussion, is that of the concept of entropy in thermodynamics. First, entropy is not a directly observable magnitude, but is, rather, a complex function of observable magnitudes. Second, the formulae describing the regularities pertaining to entropy are not 'deduced' from the observed phenomena. The empirical support of the Second Law of thermodynamics is provided by the fact that perpetual motion has not been found in nature, in the sense of lawful or lawlike net production of work from heat for some considerable time interval. This can be generalised to saying that perpetual motion does not exist in the universe (as famously expressed by Clausius and Kelvin, see Fermi 1936). But thermodynamics tells us more, namely, that perpetual motion is impossible (in some appropriate sense of 'possible'; see Uffink 2001), and this is no longer an empirical generalisation, although it may of course explain the above generalisation. But Einstein's imprecision here is not important for our case: the reader may employ their favourite approach in the philosophy of science, as long as some sort of strong connection to the observations is maintained. Howard and Giovanelli (2019, §6), for example, describe 'principle theories' as 'individually well-confirmed, high-level empirical generalizations', and this suffices for us here.

'*Constructive theories*', by contrast, are hypotheses concerning theoretical entities, that are intended to explain the facts described by the principle theories. In Einstein's view, since the principle theories are strongly connected to the observations, the constructive theories must comply with them, rather than vice versa; this compliance is a constraint or criterion that constructive theories must satisfy. (Adhering too strongly to this constraint may impair the advancement of science; see Callender 1999).

It is important, for our argument, to notice that Einstein's distinction between principle theories and constructive theories is context dependent: for example, Howard and Giovanelli (2019) notice that in one context, e.g. the reduction of thermodynamics to mechanics, Boltzmann's statistical mechanics is a constructive theory, while in another context, i.e. Einstein's (1905c) photon hypothesis paper, Boltzmnn's notion of entropy given by the famous equation $S = k \log W$ is treated as part of a principle theory.

We do not wish, in this paper, to commit to any notion of inter-theory relations nor to any notion of explanation, and therefore we shall not argue for, nor expand any further on, Einstein's distinction. We use it in order to provide an example of what one might generally call a *deep structure* account of phenomena. We take it as by and large non-controversial that all the sciences ultimately aim to provide deep structure accounts to some extent. (We stress: *deep* structure does not entail *deepest* structure[10]). Our point with regard to Hempel's Dilemma should be applicable given one preferred notion of deep structure.

We argue in this paper that Hempel's Dilemma applies to any deep structure account, that is, in all cases where phenomena or phenomenological generalisation (where thermodynamics is the paradigmatic example) are accounted for by a hypothetical deep structure (and here statistical mechanics would be a paradigmatic example). The deep structure theories are, as stressed by Einstein (see Giovanelli 2020), constrained by the phenomenological generalisations; but this, as is well known, leaves them under-determined, and with ample room for change. This room for change is arguably part of the strength of science. But it obviously opens the door for Hempel's Dilemma. Instead of discussing this in the abstract, we shall present further details and arguments by way of discussing several cases that illustrate the generalised scope of Hempel's Dilemma, emphasising the role of deep structure accounts, and using for convenience Einstein's distinction between principle theories and constructive theories.

We repeat and stress that the reader may prefer other approaches concerning the relationship between the physical facts and the facts that are to be reduced to them, but our argument should apply for that preferred approach as well: *For any account of the relationship between physics and the facts to be reduced to it, there is a corresponding Hempelian Dilemma.*

3.2. Requirement 2: Theory Change

Hempel's Dilemma concerns only cases in which the low-level theories can change, or be replaced by other theories. While opinions vary on these matters, as we have seen above, one thing is clear: for Hempel's Dilemma to be applicable, one has to accept that the deep-structure theory *must* be changeable. The conjecture that it may change *must* have some support in contemporary science, and so *must* the unknown nature of its future replacement. Without these, the Dilemma does not apply.

Here it is important to address the view that contemporary physics is unlikely to undergo radical changes in the future, and moreover, that even if physics would change significantly in an unforeseen way, nevertheless these changes are unlikely to be relevant to the account of the mental. If this view were true, it would undermine requirement 2 of theory change, as being irrelevant to philosophy of mind altogether. However, it seems to us that there are ample historical examples which are enough to show that this view is wrong-headed.

There is a widespread belief that the renowned physicist Kelvin in his 1900 lecture (that appeared in a 1901 paper) titled 'Nineteenth Century Clouds over the Dynamical

Theory of Heat and Light', expressed the conjecture that physics will not change much after 1900. Reading his paper more carefully one finds that this is not the case; quite the contrary, Kelvin was aware that these are serious problems for his contemporary science, and the notion of 'clouds' that he used expressed this worry. (It is too much to expect that he should know how to solve these problems!) However, some contemporaries of Kelvin did think that physics will not change much. Michelson, whose famous experiment was discussed in Kelvin's (1901) paper, wrote:

> The more important fundamental laws and facts of physical science have all been discovered, and these are now so firmly established that the possibility of their ever being supplanted in consequence of new discoveries is exceedingly remote. (Michelson 1903, 23)

Michelson famously regretted these comments (see discussion and additional examples in Badash 1972). But the lesson has not been learned. Here is a famous example:

> In these pages I want to discuss the possibility that the goal of theoretical physics might be achieved in the not-too-distant future: say, by the end of the century. By this I mean that we might have a complete, consistent and unified theory of the physical interactions that would describe all possible observations. Of course, one has to be very cautious about making such predictions. … Nevertheless, we have made a lot of progress in recent years, and as I shall describe, there are some grounds for cautious optimism that we may see a complete theory within the lifetime of some of those reading these pages. (Hawking 1980)

This latter writer, too, soon changed his mind, at least to some extent:

> Some people will be very disappointed if there is not an ultimate theory that can be formulated as a finite number of principles. I used to belong to that camp, but I have changed my mind. I'm now glad that our search for understanding will never come to an end, and that we will always have the challenge of new discovery. Without it, we would stagnate. (Hawking 2002)

At present, there is a general agreement that the task of science (and specifically that of physics) is far from being completed, since there are deep problems at the foundations of all theories in science. Interestingly, however, and relevant for Hempel's Dilemma(!), there is no agreement concerning *which* are the problems at the foundations of science that are likely to lead to change, let alone their solutions and their implications. For example, it is telling that the list of such major problems by Peebles (2005) and by Smolin (2006) are very different.[11] So the fact of the matter is that we do not know which parts of contemporary physics are false and to what degree (first horn of Hempel's Dilemma); and moreover we do not have a clue as to the nature of future physics, or how close or remote it will be to any part of the contemporary theories (second horn). For example, amongst the interpretations of quantum mechanics we can find radically different theories such as hidden variables theories (e.g. Bohm 1952) and stochastic collapse theories (e.g. Ghirardi, Rimini, and Weber 1986), some of which require deep changes of the fundamental principles and symmetries of standard quantum mechanics, such as the unitary dynamics, the relativity of simultaneity and the equivalence of all reference frames. In various approaches to quantum gravity, spacetime, which is taken to be fundamental in contemporary theories, turns out to be *emergent*, and moreover, some of the *fundamental* principles of both quantum field theory as well as relativity theory will have to be revised; e.g. the commutativity of observables pertaining to spacelike separated regions required by quantum field theory may not be fundamental (for observables in the bulk in Anti-de Sitter space[12]; see Harlow 2018). This is so even though quantum field theory is assumed to hold with respect to the boundaries of the space. In this case, nonlocal signalling is in principle possible and absolute time order emerges (in the bulk).

Since even the structure of spacetime itself in fundamental physics will presumably undergo immensely radical changes, which are not yet fully understood, it seems to us that the view that such changes are unlikely to be relevant to philosophy of mind is likewise unfounded and expresses hope rather than known facts. In von Neumann's (1932) standard formulation of quantum mechanics, which is our best contemporary fundamental framework of quantum field theory, the *mental* is an indispensable part of the physical theory: that is, the 'observer's *mental* states' at the end of the measurement chain of interactions introduced explicitly by von Neumann are the ultimate empirical justification as well theoretical justification (according to von Neumann 1932) for the so-called projection postulate, where the fact of the matter is that without the projection postulate standard quantum mechanics has no empirical content whatsoever. This is just part of the so-called quantum measurement problem. Some of the interpretations of quantum mechanics propose the conjecture that the collapse of the quantum state is triggered by the mind (e.g. Chalmers and McQueen, forthcoming), and that quantum superpositions in the brain may be relevant for understanding the nature of the mind (e.g. Hameroff and Penrose 2014). So regardless of whether one seeks a pure physicalist or a straightforward dualist account of the observer in quantum mechanics, given the present state of the art, it seems quite immature to say that a deeper understanding of quantum mechanics is unlikely to be relevant to philosophy of mind. The truth is that we don't really know.

Physics is by no means the only theory that underwent radical changes in the twentieth century, and moreover that is conjectured to change further in the future. Yet, the more the entities and properties (etc.) of a given theory are close to (what many take to be) directly observable (i.e. non-theory-laden facts, if there are such facts), the more people may reasonably tend to conjecture that these theories will not undergo substantial change, and so the less these theories are subject to Hempel's Dilemma, given requirement 2. However, one important lesson from the history of science is that also such relatively robust theories are subject to radical change: consider the history of the theory of heredity (we do not use the term 'genetics' on purpose here!), or of the origin of species (we do not use the term 'evolution' on purpose here!), or of the very concept of life, or of the nature of the heart, or of the brain, and so on. (Compare Hoefer 2018 and Hoefer and Marti 2020, who distinguish between fundamental theories that are more likely to change and other theories, or parts of theories, in which, as they write, we have more 'handles' on reality and hence are less likely (in their view) to change.). In all of these contexts it is not always easy to disentangle the observations from the theories to which we have become accustomed and with which we trust our personal well being; the lesson from the history of science, as well as from the philosophy of science, is quite clear on this point.

We are not committed to any specific approach concerning theory change and concerning the conditions under which theories ought to change or be replaced, nor the criteria by which we should say that a theory has changed or has been replaced. *Our claim is that for any account of theory change, there is a corresponding Hempelian Dilemma.*

Finally, two remarks: (i) Note that not only scientific deep-structure theories are subject to change, but also some non-scientific ones. As an illustration, consider the appearance of a new prophet with an updated message concerning the nature of the universe. (ii) To the extent that Hempel's Dilemma concerns the account of *high-level theories* (rather than primitive facts or phenomena) in terms of *low-level or fundamental* ones, we here refer mainly to changes in the latter, although the former may undergo change as well, and this will give rise to a further expansion of the applicability of the Dilemma (see Section 4.3).

3.3. Generalised Hempelian Dilemma

We propose that requirement 1 and requirement 2 are not only characteristic of the cases for which Hempel's Dilemma has been applied in the literature, namely, physicalism, and in particular mind–body physicalism, but they give rise to a generalised dilemma, which is this:

Generalised Hempelian Dilemma. For every case in which requirement 1 and requirement 2 are satisfied, a Hempelian Dilemma applies with exactly the same structure as in Hemepl's original Dilemma.

In other words, give us your favourite account of deep-structure theory and of theory change, and we shall give you back a suitable Hempelian Dilemma.

Given this generalisation, the Dilemma's domain of application becomes quite wide, and includes theories that compete with physicalism in a variety of domains, including the context of the mind–body problem. We give examples in Section 4. Since our Generalised Hempelian Dilemma equally challenges physicalism and its rivals, for example mind–body dualism, it is no longer a threat to physicalism *as opposed to its rivals*, and therefore ceases to be an argument against physicalism *as opposed to its rivals*. It is a threat to physicalism as well as its rivals, in the sense that it becomes a general sceptical argument against any theory with a changeable deep-structure account of the phenomena (or of our experience and its generalisation).

Perhaps the approach to Hempel's Dilemma that may seem at first sight to overcome this difficulty is the stance view (see e.g. Ney 2008b). This approach is *non-cognitivist*. It says that a commitment to theories of physics (or science more generally) is not a commitment to a truth value, but is rather a commitment (in the sense of *promise*) to act in a certain way, namely, *as if* the contemporary theories of science (where 'contemporary' is indexical) were true, but without *really* believing that they are (for 'really' here, see Fine 1984). Whereas the other attempts to solve Hempel's Dilemma, mentioned in Section 2, focus on the case of physicalism and in particular mind–body physicalism, and do not apply for all of the extended domain of our generalised Dilemma (see examples below), this commitment approach suits the generalised Dilemma as well. Of course, this view has its demerits as well. It seems to us that this approach repeats the Dilemma in some sense although it is intended to resolve it: while one is committed to applying contemporary physics, be its content as it may, this commitment is problematic in case one is aware of the flaws and lacunae in contemporary physics and in this sense one's commitment to applying contemporary physics are weakened. Moreover, as we already said, in contemporary quantum mechanics, for example, the mental is part of the physical theory' e.g. in von Neumann's (1932) standard formulation of the theory (as in fact reflected by the measurement problem in quantum mechanics), and this fact significantly amplifies the Dilemma even if one adopts this approach.[13] Our main task in this paper is to point out the expanded scope of the Dilemma and the implications of this expansion. We now turn to illustrate our generalised Hempelian Dilemma for some important cases.

4. Hempel's Dilemma for Psychophysical Dualism

In contemporary literature Hempel's Dilemma is employed mainly as an argument against physicalism concerning the mind–body problem, where the 'mental' is to be reduced to the 'physical', and the main question is what those are. In this section, we will see that the Dilemma applies to mind–body dualism as well, since both requirements (i.e. of deep structure changeable theories) hold for it.

4.1. Explicit Psychophysical Dualism

Psychophysical dualism consists of three elements: (i) a deep structure theory about the material (e.g. about the facts described in the physics or physiology of the brain), (ii) a deep structure theory about the mental (e.g. about the facts described in observations and theories of psychology), and (iii) something that accounts for the correlations between the facts in these two domains (*if* they are taken to be correlated, and in dualist theories they usually are; if they are not we are left with (i) and (ii) only). Dualism is subject to Hempel's Dilemma on all three fronts.

It is convenient to begin by asking: What is the mental, about which dualism is to provide an account?

The via negativa route for characterising the mental is that it is something non-physical. Often the term *via negativa* is used in the reverse direction, namely, to characterise the physical as anything non-mental, as we saw in Section 2. Is the idea symmetrical? If it is, we may be able to characterise the mental as non-physical. But then to characterise the mental we shall need first to characterise the physical. And so, on this account, dualism is subject to Hempel's Dilemma via the criticism concerning the notion of 'physical', so that in the context of Hempel's Dilemma physicalism and dualism rise and fall together. But the via negativa may not work for other reasons as well, predominantly because 'physical' and 'mental' are not necessarily complementary terms. To see why consider the following.

Montero (1999) shows that 'physical' is *not* the same as 'non-mental'. She proposes to *replace* the mental-physical distinction with the mental-nonmental distinction, writing as follows:

> One advantage, however, is that, arguably, we do have a grasp of one side of the divide - that is, the mental side. So, perhaps, rather than worrying about whether the mind is fundamentally physical, we should be concerned with whether the mind is fundamentally non-mental. And this, I should mention, is a concern that has little to do with what current physics, future physics, or a final physics says about the world. (Montero 1999, 194)

But do we indeed have a direct grasp on the mental side? Let us see.

The direct account of the mental realm requires that there be an account of the mental irrespective of the physical–chemical-biological (apart from the account of the mental-physical-etc. correlations). Here are four options for a direct account.

The first attempt might be a *no-theory* approach to what counts as mental. One might begin by considering paradigmatic cases of the mental and generalising a bit from them without explicitly or implicitly formulating a theory. Montero (1999, 183–184) addresses the question whether the '*physical*' can be characterised in this way on the basis of paradigmatic cases like tables and rocks. She writes:

> While there appears to be something correct about the claim that we can identify central cases of being physical - what could better exemplify the physical than things like rocks and trees (except, perhaps, quarks and leptons)? - there is an extra wrinkle: rocks and trees (as well as quarks and leptons) are identified as central cases only on the assumption that idealism is false. For there is not much point in arguing about whether the mind is physical if our central examples of physical entities are entities composed entirely of sense-data. (And to say that rocks are a central example of physical objects or that the properties of rocks are central examples of physical properties only if rocks are physical, obviously does not provide us with a useful clarification). (Montero 1999, 184; our underline)

We can apply Montero's argument to the case of the 'mental': to identify 'pain', for example, or 'feeling the angst of post-industrial man under late capitalism' (Searle's example) as paradigmatic cases of the 'mental' is to assume that the mental is, for example, not physical; otherwise those are cases of the physical. So, following Montero's (1999) argument, relying on such *intuitively* paradigmatic cases is not helpful, and upon reflection we cannot claim to know what is the mental (or indeed whether some fact is mental or nonmental!) *without relying on some theory*.[14] For this reason, it seems to us that Montero's (1999, 194) proposal for replacing the mental-physical distinction with mental-nonmental is wanting.[15] We know what is mental only on the assumption that physicalism (or some other deep structure theory) is false; that is, if we deny in advance that the mental is, for example, physical! After all, according to physicalism, we don't actually know what the mental is, this is the whole point of physicalism as a specific deep-structure approach to the mind. In particular, this point holds with respect to reductive identity physicalism, but it may also hold for non-reductive physicalism. One may say that pain is a paradigmatic case of the mental only if one already rejects (for example) physicalism. Note that we do not really know whether an octopus feels pain even though it does exhibit pain behaviour. Also note that there are views in the literature (e.g. Dennett and Kinsbourne 1992) that are sceptical even about paradigmatic cases of a first-person beliefs concerning one's own mental states.

So how else can the mental be characterised? Three other options for direct accounts spring the mind; for the next two, we need to make use of Einstein's (1919) distinction principle vs. constructive *theories*.

A second option for a direct account that doesn't go via physics-chemistry-biology is to treat the mental realm in terms of a *principle* theory, for which there is *no* constructive theory. On this option, the mental is amenable to empirical generalisation, and this is all that there is to it: there is no deep structure in the world although accounting for the mental requires some principle theory. Indeed, some thinkers describe the mind this way (although we are not sure that they would adhere to this characterisation of a principle theory or of a no-theory approach). For example:

> Consciousness is not an explanatory construct, postulated to help explain behavior or events in the world. Rather, it is a brute explanandum, a phenomenon in its own right that is in need of explanation. (Chalmers 1996, 188)

> Where consciousness is concerned the existence of the appearance is the reality. If it seems to me exactly as if I am having conscious experiences, then I am having conscious experiences. (Searle 1997, 112)

However, as we know from the philosophy of science and other branches of inquiry, many and arguably all statements about phenomena are *theory laden* (see considerations and references in Bogen 2017). This includes statements about the mental realm, and in particular, statements based on first-person reports about mental experience.

> An introspective judgment is just an instance of an acquired habit of conceptual response to one's internal states, and the integrity of any particular response is always contingent on the integrity of the acquired conceptual framework (theory) in which the response is framed. (Churchland 1996, 70)

> It is not something too obvious to need support, however appealing it may at first seem. It is a first shaky step toward a philosophical theory. I offer a rival theory. (Dennett 2018)

(See further discussion in Shenker 2020.)

A third option of a direct account that doesn't go via physics-chemistry-biology is to provide an *alternative* constructive *deep-structure* account of the mental. We reiterate that this deep structure has to be non-physical (where the latter term is *not* to be understood in the via negativa terms). If a dualist theory is to have some *substantial* content, it needs to provide details about the mental realm: a dualist theory with substantial content is *not* a version of brute *mysterianism*, it *isn't* a statement concerning a 'black box'; labelling the black box 'ectoplasm', for example, doesn't count as a deep structure in the sense we are talking about here, unless details about the nature and behaviour of the ectoplasm are provided (see Robinson 2020). If dualism *is* just a black box theory, then we grant that it is *not* subject to Hempel's Dilemma, but in this case we find it so poor as to be unworthy of examination; in this case we find it a blunt expression of not willing to make the effort needed to understand the mental realm. (By 'black box' theory we do *not* include *a priori* conceptual theories of the mental; see below.) If the dualist theory of the mental is not a black box theory, then it posits a *non-physical deep-structure* theory about the mental realm, satisfying Requirement 1 for the applicability of Hempel's Dilemma.

To be subject to Hempel's Dilemma, this non-physical deep-structure theory needs to satisfy Requirement 2 as well, that is, to be subject to possible change, as a matter of principle. The changeability of the non-physical deep-structure theory depends on the methodology or the kind of justification provided for its endorsement. For example, if the methodology involves relying on revelation, and there is in-principle-openness to the possibility of some future revelation that may reveal novel and/or details about the mental realm, then this non-physical deep-structure account is subject to Hempel's Dilemma. If, however, the account of the mental is taken to be based on *unchangeable* foundations, then this theory will not be subject to Hempel's Dilemma.

If the methodology that gives rise to the non-physical deep-structure theory of the mental allows for theory change, then its situation is quite similar to that of physics in mind–body physicalism (or in the physical account of other high-level theories). Contemporary dualist theories do not provide any empirical predictions over and above those provided by empirical generalisations of the phenomena, or by the deep structure physicalist theories. This seems to indicate that contemporary non-physical deep-structure theories of the mental are either *very partial* or *false*—similarly to what is to be said about physics in this context. This is the first horn of Hempel's Dilemma now applied to dualism. The second horn is that those un-successful dualist theories may be replaced by new and better ones, of which we have no idea at the present.

A fourth option of a direct account of the mental is based on allegedly *a priori* considerations. In the past, some theories of *matter* were based on considerations that seemed obvious or true in virtue of the very concept of matter (e.g. from Descartes's conception of matter to Lewis's Humean supervenience view; see Lewis 1986, and a critical account in Loewer, forthcoming; Hemmo and Shenker, forthcoming). To the extent that *a priori* considerations are subject to change, this kind of account can satisfy Requirement 2 for Hempel's Dilemma, and since it also satisfies Requirement 1, it is subject to the Dilemma. Today, the theories of physics describe phenomena that famously challenge our intuitions and our ideas about what is *a priori* true (for example concerning locality, quantum superposition and entanglement, and the nature of space and time). Theories about the mental are to a large extent based on *a priori* considerations, such as for example conceptual analyses of the 'subjective' and 'objective'; but even here experimental results pose serious challenges to our intuitions (for example concerning the way in which our sense of 'self' and the contents of the 'present experience' are constructed and influenced, e.g. Dennett and Kinsbourne 1992; Graziano 2019). Of course, if the deep-structure nature of the mental is known in a way that is not subject to change, then this account will not be subject to the Dilemma; but this view is subject to other serious objections.

Similar considerations apply for (iii) above: the account of the correlations between the mental and the physical (in case the material realm and the mental realm are taken to be correlated). Chalmers (1996, 2013) advances an explicit dualist view. In his view, the physical and the mental are not just accidentally correlated: there are 'fundamental principles connecting physical processes to conscious experience' (Chalmers 2013, 34). The metaphysical status of these principles is a matter yet to be investigated. First, the identification of these principles between the physical and the non-physical is subject to identifying the physical and the mental, that is, to the solution for (i) and (ii) above. But here the worry that gave rise to Hempel's Dilemma, concerning deep structure theory change arises for (iii) above, no less than in the context of physicalism.

In sum, Hempel's Dilemma is a challenge to mind–body dualism no less than to physicalism, and therefore it cannot be used as an argument favouring one approach over the other.

4.2. Non-reductive Physicalism

Non-reductive physicalism (NRP) is the idea that the facts (properties, kinds, etc.) described in some so-called special science, such as biology or psychology (often called 'high-level' facts), are not identical with physical kinds (or 'low-level facts'), but only supervene on them, allowing (as a matter of principle) for multiple-realisability (even if as a matter of contingent fact no case of multiple realisation is found; see for example Polger, Shapiro, and Lawrence 2016; Maimon and Hemmo, forthcoming). NRP has become very dominant in contemporary philosophy as well as science (see the definition of 'physicalism' in e.g. Dowell 2006a, 2006b; Stoljar 2017), since it offers a combination that many want to endorse: on the one hand, accepting naturalism and even physicalism,[16] through the commitment of supervenience of the high-level facts on the low-level facts, and on the other hand—and seemingly coherently!—accepting that there are facts that are *not* reducible to physics, thus giving place to prevalent intuitions of two sorts: first, that the special sciences are autonomous from physics (e.g. Fodor 1974, 1997), and second, that there is something about the mental that is not entirely physical (e.g. Putnam 1967; Davidson 1970). Since supervenience is the *logical* relations between the high-level facts and the low-level facts, there is room for explaining how this relation comes about, and here there is a variety of *metaphysical* approaches, e.g. realisation, grounding, emergence and perhaps more; we do not address the (important) differences between them.

Hempel's Dilemma applies to NRP with respect to the low-level facts and with respect to the high-level facts, in different ways. Begin with the low-level facts. NRP is subject to Hempel's Dilemma with respect to its physical basis, on which the high-level facts *supervene*: difference in the physical basis entails difference in the details of the metaphysical relation with the supervening facts (see e.g. Montero 1999). Of course, one of the appealing features of NRP is the thought that because of the *multiple-realizability* of the high-level facts by the low-level ones, the connection to the physical basis is weaker than in reductive physicalism: the physical basis which is supposed to account for the higher-level facts is only one such basis, and hence the question posed by Hempel's Dilemma (which physical basis, according to contemporary or future physics?) may be seen as a bit less important for those high-level facts. So let us look at the high-level facts in NRP and see whether they may be subject to Hempel's Dilemma over and above the low-level physics.

Consider the proverbial Demon of Laplace, who knows all the facts at the low-level of physics and can make all the calculations concerning its evolution. And consider three tokens (a, b, and c) that belong to three physical kinds (A, B, and C respectively), such that A and B fall under the high-level kind M but C doesn't. *Given* this partition into kinds with its logical supervenience structure, if we are given a token with its physical description, Laplace's Demon can deduce whether or not it belongs to M. In this sense, the low-level facts determine the high-level facts. The *partition* of physical kinds and tokens to *high-level* kinds is a *fact* about the world, that is epistemically accessible even to finite creatures like us, so that we can make meaningful statements about high-level facts, statements with reference and truth value.

Laplace's Demon can know all the physical properties, and can therefore partition the tokens into physical kinds, i.e. into equivalence-sets in which all the tokens share the same physical property. Given a token, the Demon can know all the physical kinds to which it belongs. But can the Demon also partition the tokens into high-level kinds, and know which high-level properties a given token has? This depends on the sort of fact that the partition to high-level kinds is. What fixes this partition? And why is the case that the partition satisfies supervenience? Do the metaphysical theories concerning realisation, grounding, emergence, etc. provide an answer to these questions? (We shall not go into the details of these different metaphysical approaches.) Let's explore these questions a bit more.

If reductive type-type identity physicalism is true, so that M is just another name of some given physical kind, then of course the Demon can easily know for every token whether or not it is a case of M. But we are interested in NRP, and therefore *by assumption* this is not the case. To rule out reductive physicalism and stick to NRP we must assume that a and b don't share *any* physical property, or at least don't share any physical property that is somehow *relevant* to their having the high-level property M; we say that a and b are *physically heterogeneous*. Presumably, c (which isn't M) is also heterogeneous relative to a and to b (both of which are cases of M). Let us assume, for simplicity, that the tokens a, b, and c are all mutually heterogeneous to the same degree: If a and b were physically more similar to each other than both are to c then we might have had here a case of reduction to physics (since similarity is coarse-grained identity); but we are interested in genuine NRP cases.

What fact, then, can make it the case that—among the three tokens that are all mutually physically heterogeneous to the same degree—a and b are M but c isn't? Again, this fact fixes the high-level facts, which are part of the ontology and are epistemically accessible to us, and therefore they are part of the ontology, which is present *in each and every token*, enabling us to describe the high-level kinds to which every given token belongs. What kind of facts are they? And what are the consequences for Hempel's Dilemma? There are various proposals in the literature concerning the nature of the facts that fix the high-level properties of the tokens. We will not examine all of them here (this is beyond the scope of this paper), and we shall only comment on what is relevant directly for Hempel's Dilemma and the conditions for its applicability (as described in Section 2).

Whether or not the facts that fix the partition of tokens (and of physical kinds) to high-level kinds, are subject to Hempel's Dilemma, depends on whether or not those are theoretical facts that allow for changeable deep structure accounts. Let us very briefly mention two popular ideas concerning the way that the high-level facts come about, just to illustrate the point.

One prevalent view concerning the way in which high-level kinds are formed is functionalism. The idea is that—in our above example—although a and b are physically heterogeneous to the same degree that c is heterogeneous with respect to them—a and b share the same functional role, that is not shared by c; and this functional role is associated with M (or implements M, etc.). The idea is that a and b are each a state within a sequence called Sa or Sb respectively (e.g. a causal or a computational sequence), such that Sa and Sb are of (or implement etc.) the same function (causal or computational). But what makes the two sequences Sa and Sb members of the same *functional kind*? Presumably, 'implementing function M' is not a physical property, for then we would have a case of reduction to physics, whereas we are interested in genuine NRP. It turns out that functionalism doesn't solve the problem of explaining how high-level partitions come about, but repeats it (for more details see author reference).

Another prevalent view concerning the partition of physical kinds to high-level kinds is that this is a brute fact that is not explained, nor does it call for explanation; (see e.g. Vintiadis and Mekios 2018). These brute facts exist and are epistemically accessible to us, so that we can make sentences in the special sciences that are about the high- level facts, sentences that have meaning and reference and truth value, and we can determine whether a particular high-level fact obtains in particular (token) cases. Concerning these facts, the main point of the *brute facts* approach is that there is *no* deep structure: they are fundamental (or basic or elementary, etc.) facts, that—as a matter of principle!—will not undergo theoretical change and aren't and needn't be explained. Since brute facts don't satisfy the conditions for the applicability of Hempel's Dilemma, they are not subject to it.[17]

There are other proposals in the literature concerning the formation of high-level kinds that supervene on physical kinds, for example: the high-level kinds appear in the regularities described by the laws of the special sciences (see e.g. Fodor 1974; Polger, Shapiro, and Lawrence 2016); or they are structures that emerge from the fundamental physical level (e.g. Wallace 2012; see overview of the notion of emergence in O'Connor 2020). We think—but will not argue for this here, since this is not the focus of this paper—that all the explanations boil down, on deeper examination, to two kinds of explanations: either accounts by *reducing* the high-level facts to the low-level physical facts; or accounts that bring into the picture *non-physical* facts that obtain *as part of each token*, which is arguably a case of ontological dualism (see Shenker 2017; Hemmo and Shenker 2019, 2020). In any case, whether or not a theory or an account or a fact is subject to Hempel's Dilemma and its implications depends only on whether Requirement 1 and 2, of a changeable deep structure account, are satisfied.

4.3. A Remark on Hempel's Dilemma for Non-fundamental Theories

Hempel's Dilemma applies to sentences of the form 'A is B' in which the term 'is' is understood as any kind of reduction that the reader finds relevant in this context. In Sections 2 and 3, we addressed an expansion of the B term, to anything that satisfies Requirement 1 and 2. In this section, we wish to further expand the domain of applicability of Hempel's Dilemma, concerning the B term and then the A term.

4.3.1. Expanding the Range of Reducing Theories: Not Only Fundamental Ones

We begin with the B term. Above we focused on cases where the reducing theory is fundamental, whether it is physics or some other theory.

> That some x is metaphysically fundamental means, roughly, that x is a privileged entity, or set of entities, playing a special role in determining the structure of reality. (Morganti 2020)

A Hempelian Dilemma can also arise with respect to non-fundamental theories. For example, if the mental is understood in terms of the physiology of the brain:

(1) The mental is biological.

Hempel's Dilemma arises in case the current theories of physiology are false and the future ones unknown, *regardless* of whether or not the physiological is fundamentally physical. The theory of physiology might change even without physics changing. The intuition might be that as the theoretical entities, properties and laws are nearer, in some appropriate sense of the term, to pre-theoretical ones (to the extent that the theory ladenness is not substantially misleading), the chances for significant theory change is smaller. But the history of science teaches us that this is not the case. The considerations here which are well-known in the philosophy and history of science, are put to work in the particular context in which we are now interested.

To the extent that Hempel's Dilemma applies to non-fundamental theories, it is applicable in case there is no fundamental theory or realm at all (see Morganti 2020).

In what follows we shall stick to standard examples, in which the reducing theory is fundamental physics, but everything we say applies also to non-fundamental theories (as in Section 4.3) and to non-physical theories (as in Sections 4.1 and arguably 4.2).

4.3.2. Expanding the Range of Reduced Theories or Phenomena

Hempel's Dilemma is most frequently applied to reducing the mental to some deep-structure domain, such as in (1) above Sec 4.1, but it applies equally well to the reduction of other so-called 'high-level' domains, for example in

(2) The biological is physical

In the original formulation by Hempel and by Crane and Mellor (see Section 1) there is no mention of the mental realm, but we think it is clear from the context that this Dilemma is applicable to any suitable reduction of any 'high-level' domain. Perhaps the most important statement of this form is this:

(3) The facts described by thermodynamics are facts of mechanics.

This statement is perhaps the most important example in science of a project of reduction (in whatever sense of 'reduction'). Kripke (1980) took his example 'heat is molecular motion' from this context of the ongoing project of reducing thermodynamics to mechanics (we discuss Kripke's ideas in Section 5). In the present section, we focus on the reduced principle theory of thermodynamics, rather than on its reducing theory statistical mechanics. Einstein (1919) took thermodynamics to be the paradigm of a principle theory, and later wrote:

> [Classical thermodynamics] is the only theory of universal content concerning which I am convinced that, within the framework of the applicability of its basic concepts, it will never be overthrown. (Einstein 1970, 33)

Eddington famously added:

> The law that entropy always increases, – the second law of thermodynamics – holds, I think, the supreme position among the laws of Nature … . [I]f your theory is found to be against the second law of thermodynamics I can give you no hope; there is nothing for it but to collapse in deepest humiliation. (Eddington 1935, 81)

These thinkers thought that thermodynamics is unshakeable (see Hemmo and Shenker 2012, 2016 to the contrary) because they took it to describe no more than generalised observations. It is important to notice that while thermodynamics is very strongly supported by experience, its entities and properties and laws are theoretical ones, far richer in their contents than mere generalisation from experience. As in Section 3.1, we stress here that the history of science teaches us that all theories can change.

'High-level' theories like zoology and physiology and geology have substantially changed, and we propose that this changeability may bring about a variant on Hempel's Dilemma, in so far as the A terms are theoretic or theory laden.

5. The Role of Identity Statements in Hempel's Dilemma

Hempel's Dilemma applies to statements that seem to be identity statements, especially in the case of physicalism. In this section, we (very briefly) address the nature of identity statements and distinguish between those for which the Dilemma applies and those for which it doesn't. The key for this (in)applicability is—as everywhere in our

Generalised Hempelian Dilemma—the satisfaction of Requirement 1 and 2. Let us begin with the sentences for which Hempel's Dilemma is most frequently applied:

(a) Everything is physical,

and a sentence that follows from it:

(b) The mental is physical.

As we saw in Section 4, although sentences (a) and (b) are those for which the Dilemma is most often applied, its proper domain of application is much wider with respect to both the reducing and the reduced domains, as long as the reducing theory is a changeable deep-structure theory, that satisfies Requirement 1 and 2. In the present section, we shall compare sentences like (a) and (b) with another sort of identity statements, that in our discussion have an important role in understanding the scope of Hempel's Dilemma, but as we shall argue below they are not themselves subject to this Dilemma.

Assuming that 'pain' is a particular case of 'the mental' that appears in (b), and that 'C-fibres firing' (following Kripke's (1980) famous example[18]) is a particular case of 'the physical' in (b), the following statement is a special case of (b) as well as of (a):

(c) Pain is C-fibers firing.

We think (following Kripke 1980) about sentence (c) as an *identity* statement. In our analysis, we do not follow Kripke in all the details and some of the following account is our own view of the matter.[19] The sentence (c) is *not entailed* by (a) or (b), because it may turn out that although everything including the mental is physical, pain is identical with some other kind of brain states or something else altogether.

Is (c) true? Kripke (1980) thinks that (c) is false even if one thinks of 'C-fibers firing' as a placeholder for *any* physical (or biochemical, neurological, etc.) kind, and this is part of his (so-called) conceivability argument supporting mind–body dualism. As part of the argument, Kripke compares sentences like (c) with sentences like (d):

(d) Heat is the motion of molecules,

where by 'motion of molecules' Kripke means (we think) some specific function (or macrovariable) of the microstate of the system to which the property we call 'heat' is assigned; for example, in the case of an ideal gas in equilibrium, 'heat' is (or means) average kinetic energy. Here, our aim is to analyse the nature of sentences like (c) and (d) in order to connect them to our Generalised Hempelian Dilemma. Like (c), (d) is an identity statement. Following Kripke, we think of (d) as an identity sentence that was discovered scientifically and is part of contemporary physics. (d) is not entailed by (a), because it may turn out that although everything is physical, contemporary physics is wrong.

To see better what is going on here, consider the following sentence, which in the past was considered true, but is now believed to be false:

(e) Heat is caloric.

(an idea mentioned by Hempel 1980, as quoted in Section 1). According to current science (e) is false, whereas (d) may be true.

Before proceeding to explore the status of (d) and (e) and what they can teach us about Hempel's Dilemma, we need to *digress (for one paragraph)* to say something about the difference between (a) and (b) on the one hand and (d) and (e) on the other.[20] The following question arises. Assuming that we can reasonably say (as we just did!) that according to contemporary science (e) is false but (d) may be true, why can't we assign truth values (true or false) to (a) and (b) in the same way? Our answer is that this difference between these statements is precisely the main point of Hempel's Dilemma. Let's see why, beginning with (d) and (e). Since 'caloric' and 'molecular motion' are theoretical terms, one may say that their meaning is only relative to the theory in which they are embedded, and so deeming statement (e) false essentially means rejecting the relevant theory, and deeming (d) possibly true means that the theory in which it appears is acceptable (at least tentatively). *Given* the relevant theories, and within their context, (d) and (e) receive reasonably definite meanings. This is why (d) and (e) are not subject to Hempel's Dilemma. By contrast, Hempel's Dilemma is the claim that the term 'physical' in (a) and in (b) does not have a definite meaning. Whether or not this is the case is the debate around Hempel's Dilemma: The attempts to solve Hempel's Dilemma try to provide a definite meaning to the term 'physical', e.g. by saying that 'physics' is the contemporary theories; or that future physics will be sufficiently similar to the contemporary one; or that 'physics' is any theory that is deemed best at the time in which one is working; etc. These attempts at making the term 'physical' in (a) and (b) definite have met with criticisms. In our view these criticisms are by and large on the right track, and Hempel and Crane and Mellor and their followers are correct in accepting that indeed the term 'physical' in (a) and (b) is not definite. We *embrace* this conclusion, and *generalize* it, and point out that it is an expression of a general anti-scientific sceptical argument, with which we need to live (or if the reader prefers, for which there is some acceptable response), treating science with the suitable caution and humbleness. Here we end this digression, and from now on we assume that (d) and (e) can be ascribed meaning and truth value such that (e) is false and (d) is possibly true, and our focus will be on arguing for the claim that while (a) and (b) are subject to Hempel's Dilemma (d) and (e) are not.

Returning to developing our argument, let us now compare (d) and (e) with the following sentence:

(f) Heat is physical.

The term 'physical' in (f) refers to a changeable deep-structure account of heat, unlike the terms 'caloric' and 'molecular motion': those would be changeable only if we take them to be mere placeholders; but we take it that this is not how these terms are normally understood in (d) and (e): they are normally understood to have specific meanings and references (even if they are false and even if they are understood as fictional terms). Because of this, (f) is subject to our Hempelian Dilemma, much like the statement (a) from which it can be derived; whereas (d) and (e) are not directly subject to Hempel's Dilemma, since they do not satisfy Requirement 2 of changeability. The same reasoning holds with respect to (c) which is therefore not directly subject to Hempel's Dilemma.

Kripke (1980) brings the example of (d) in order to compare its structure and the relation between 'heat' and 'the motion of molecules' with sentences like © ('Pain is C-fibres firing'). Kripke's aim is to argue that whereas (d) may be true, (c) is certainly false. Emphasising the difference between these statements, Kripke (1980) brought to our attention that (d) might be understood *not* as an identity statement but as referring *implicitly* to the relationship between two systems: there is a feature of system S (an aspect, a property, etc.) that we call 'heat', and this feature casually[21] brings about in another system U, which is typically us (humans), a certain kind of (mental) state, that we may call (for lack of a better expression) 'a sensation of heat', and following scientific investigation we discover that the feature of system S that causes in U the sensation of heat is the motion of the molecules of S. And so, (d) can be understood in two ways:

First: If 'heat' is the name of *the sensation of U*, then a replacement of (d) that more completely describes what is going on here would be something like this:

(h) (1) Heat is the name of the sensation of U; and (2) Heat (that is, U's sensation) is caused by the motion of molecules of S.

Second: if 'heat' is to be understood as the name of the thing *in S* that brings about the sensation of heat in U, then a more complete phrase should be:

(j) (2) Heat is the name of the motion of molecules of S; and (2) Heat (that is, S's motion of molecules) causes a certain sensation in U (which for lack of a better name we call U's sensation 'the sensation of heat').

Compare: (h') 'Hesperus is the name of a certain *mental* state, which is a picture *in our mind* that has the shape of a star in the evening', vs. (j') 'Hesperus is the name of the evening star, out there as it were'. We then say: Hesperus in the sense of (h') is caused by Hesperus in the sense of (j') (but if we are dreaming then this causal statement may be false).

Why speak in terms of names here? Why not stick to a simpler identity sentence as in (d)? What does (j) add over and above (d)? To see this let us consider the notion of 'heat', and distinguish between several ways of understanding it.

One of the things that thermodynamics says exists in the world is heat: according to thermodynamics there is something called 'heat', that has a certain quantity, and is transferred in a certain way from one place to another. Thermodynamics doesn't say much, however, about the nature of that thing, so that in the past one idea was that the term 'heat' refers to a fluid that flows between the particles, and was called 'caloric'. We could say, in that context, that people thought that (e) was true: 'heat is caloric'. They thought that either

(k) The sensation of heat in U (a sensation that we call Heat) is caused by caloric in S.

Or

(l) Heat is the name of caloric in S.

We are bringing this historical case in order to clarify the status of 'motion of molecules' in (d) and its relation to the thermodynamic notion of heat. In terms of structure, 'caloric' in (k) and (l) can be treated as a placeholder, filled today by what current physics thinks is a better theory. More generally, both 'caloric' and 'motion of molecules' can be seen as placeholders, which is the X in the following sentences:

(m) The sensation of heat in U (a sensation that we call Heat) is caused by X in S.

(n) 'Heat' is the name of X in S.

Before we turn to see how Hempel's Dilemma is significant here, let us remark that thermodynamics tells us something about the quantity and behaviour of a certain feature *of* system S that it calls 'heat', without telling us *what* this 'heat' *is*. This is because thermodynamics is a principle theory, not a deep-structure constructive theory. As far as thermodynamics is concerned, it can only say something like this, using the templates (m) and (n):

(o) The sensation of heat in U (a sensation that we call Heat) is caused by heat in S.

(p) 'Heat' is the name of heat in S.

But (o) and (p) are much less informative (if they are informative at all) than (k) and (l) or their contemporary counterparts which are part of deep-structure constructive theories.

Let us return to Hempel's Dilemma. It turns out according to the above analysis that identity statements like 'heat is caloric' and 'heat is molecular motion' are *not* subject to the generalised Hempelian Dilemma, despite the fact that their truth and meaning depends on a theory. The reason is that in these cases the theory is fixed! That is, one may think that the theory is false, but nevertheless, the meaning of these sentences remains fixed although it depends on the (false) theory. By contrast, statements of the more general kind such as 'everything is physical' (or everything is nonmental) are subject to the generalised Dilemma precisely because they are not confined to a *specific known* theory—regardless of whether the theory is taken to be true or false.

Is (a) ('Everything is physical') true? Assuming that (d) ('Heat is the motion of molecules') is false (given that contemporary physics is false, and as we now know that (e) is), and assuming that this is a typical example of the application of all the particular applications of the physicalist hypothesis, the truth value of (a) cannot be decided neither now nor in the future (since the meaning or reference of future physics is unclear). This is the significant consequence of Hempel's Dilemma. The same reasoning applies to (f) ('Heat is physical'). And since our example here can be easily generalised, it illustrates how Hempel's Dilemma is general in the sense that it is applicable to any deep-structure but changeable account of our experience as well as of the kinds that appear in the special sciences (such as heat in thermodynamics). In this sense, Hempel's Dilemma expresses a general sceptical stance concerning deep structure changeable theories.

6. Conclusion

Hemple's Dilemma is seen, following the well known argument by Crane and Mellor (1990), as a challenge to physicalism (see also e.g. Stoljar 2010). It is from this perspective that it is addressed in the literature what are the attempt is to show that physicalism can be defended against this Dilemma. We mentioned in the introduction some of the ways in which this is proposed to be done.

In this paper we do not argue that Hempel's Dilemma is not a threat to the theory, or metaphysical framework, to which it is applied: *we leave open* the question of whether or not any of the proposed ways to defend physicalism against Hempel's Dilemma works. Instead, our point is that, if Hempel's Dilemma is a threat at all, then the domain to which this threat applies is much wider than physicalism, and includes all changeable deep-structure accounts of phenomena, not only physical (or physicalist) accounts. To clarify the notion of deep structure accounts we mentioned Einstein's notion of constructive theories, that has been proposed to explain principle theories; but this is just one example of deep structure accounts, and our argument applies to all of them.

The consequences of our argument are these. First, as a matter of principle the hypothesis of physicalism rises (if there is a solution to Hempel's Dilemma) and falls (if there is no good response to Hempel's Dilemma) together with all changeable deep structure accounts, both within science and outside of science. Second, and perhaps more generally, Hempel's Dilemma turns out to be part of a general sceptical argument against changeable theories that seek not only to generalise our experience but also to account for our experience reductively by means of some deep-structure, and should be dealt with as such.

Notes

1. Stoljar (2017), in his overview of this question, distinguishes between the 'theory-based conception', according to which to be physical is to be something that physical theory tells us about, and the 'object-based conception', according to which to be physical is to be something required by a complete account of the intrinsic nature of paradigmatic physical objects and their constituents. We address the former conception.
2. Earlier, in his (1969) paper, Hempel already remarked that (what he calls there) the 'mechanistic' thesis is obscure—and it seems to us that he would agree to our 'extended' Hempelian Dilemma presented in the next section. Hempel writes: 'Thus, we face a dilemma. If for the sake of conceptual clarity, we give the thesis of mechanism a linguistic turn, we fail to express its philosophical intent; and if we try to formulate that intent in ontological terms, or in the "material mode," the resulting statement proves to be seriously obscure and elusive. The same difficulty besets all reductionist theses that are conceived as ontological claims' (Hempel 1969, 183).
3. Much earlier Feigl (1969) discusses the problem described in the second horn of Hempel's Dilemma, and says that others are aware of it as well. He writes: 'Schlick, Carnap, and Reichenbach, who espoused this thesis [of physicalism or the unity of science], were fully aware of its conjectural and hence precarious character. Essentially it endorses a certain program for the current and future development of science towards a unitary or monistic set of explanatory premises. … As a distant goal of this program of unitary explanation, some future theoretical physics is fancied. … This thesis is, of course, not only problematic, but also inevitably vague in that such a theoretical physics may have to be very different from its current stage. All that can be said at the moment is that the 'style' of explanation might be somewhat similar to that used in the present stage of the theories of relativity, quantum mechanics, and quantum electrodynamics' (Feigl 1969, 21).
4. See for example Dowell (2006a), Pineda (2006), Ney (2008a), Stoljar (2010, Ch. 5), Bokulich (2011) and Prelević (2018).
5. Chomsky (1968, 2000, 2003) also argues that all the theories of physics from Newotonian mechanics onwards are meaningless in the sense that they carry no meaning for us, human beings. These theories are, in Chomsky's view, formalisms accompanied by instructions for application in empirical predictions. In his view, all these theories make claims that are incomprehensible by human beings; they exceed our cognitive capacities. The cognitive capacities of human beings are limited by our biological making, and the nature of the material world (as well as the nature of the mental realm) is beyond our grasp. He argues that the last theory that was comprehensible to us was Descartes's; but it makes false predictions.Dennett, in an interview, responded that while it is by and large non controversial that we are limited qua biological creatures, Chomsky provides no basis for claiming that our limitations lie in understanding physics.

6. van Fraassen (2002) applies this in characterising empiricism.
7. We talk about lower-level theories rather than fundamental theories, avoiding the need to commit to fundamentalism. See Morganti (2020).
8. There are anticipations of the distinction between principle vs. constructive theories in the nineteenth-century and early twentieth-century literature; see e.g. Bechler (1993, Ch. 3.7.10; Howard and Giovanelli 2019). Bechler (1993) argues that Einstein's idea of a principle theory is essentially connected to Poincare's conventionalism. But we set these issues aside here.
9. For another application of Einstein distinction, see Shenker (2015); Hemmo and Shenker's (2021) analyses of Davidson's (1970) anomalous monism in its terms; and compare between the structures of the arguments by Einstein (1905a) and Davidson (1970).
10. Hence we need not be committed here to there being a most-fundamental level, that is, to fundamentalism in this respect.
11. Here are some examples of such problems. One problem in contemporary physics includes the well-known incompatibility between quantum mechanics and general relativity. There are ongoing attempts to build a 'deeper structure' theory (a notion we discuss below) from which the main facts described by quantum mechanics and general relativity will emerge, but those are still work in progress and give rise to debates concerning their merits and even acceptability. Another problem in contemporary physics is the measurement problem in quantum mechanics: here some people think that the problem has been solved, but this point is under debate. A third problem is the account of the direction of time (if there is any) and—even more severely—the flow of time and the notion of the present. A fourth problem concerns the need to employ mathematical ideas in physics that some people think are dubious. There are other examples, and the reader may provide more, from physics as well as from other sciences.
12. Anti-de Sitter (AdS) space is the maximally-symmetric spacetime of constant negative curvature. Our universe (according to the best estimates) has a small positive cosmological constant. Nevertheless, the AdS space presents many of the puzzles associated with quantum gravity.
13. Also: it seems to us that the stance approach shares the merits and demerits of non-cognitivist approaches in meta-ethics (see van Roojen 2018), but we do not address this issue here.
14. An anonymous reviewer, in suggesting that we are able to know what mental is without relying on some theory, compared this to knowledge of what a square is, which is allegedly possible in a way that does not depend on a theory of squareness. People use the word 'square' in daily contexts, sometimes correctly and sometimes incorrectly. How can we say which is which? Do we not rely on theory? Famously, rich contents were added to the notions of Euclidean geometry when Hilbert offered his 20 axioms of Euclidean geometry.
15. The case of ghosts that Montero (1999) brings in invites a different sort of consideration, since here: (i) With regard to the mental vs. physical distinction, although it seems clear that ghosts are nonmental, it is harder to say whether or in what sense they are physical; (ii) Another issue concerns existence (e.g., of ghosts): one needs to consider what are the criteria for existence and whether there are different sorts of existence in the first place and what are they.
16. For an argument that physics is fundamental, see (Morganti 2020).
17. Although this is not our main subject here, let us make two very brief comments about the brute facts option. (1) We doubt that in any particular example the brute facts idea is taken seriously: we doubt whether anyone seriously thinks that the fact that my pain is *not* realised by, say, my cup of coffee, is a mere brute fact. (2) We vehemently object to the idea that brute facts accounts are even legitimately acceptable in science and philosophy, since we see them as forms of mysterianism, that merely prohibit the search for explanation and are therefore detrimental to progress.
18. This statement is false to a large extent according to contemporary science; we have learned a lot about the physiology of pain since Kripke came up with this example in the 1970s.

However this falsity doesn't matter for our present argument (especially in the context of discussing Hempel's Dilemma), and therefore we choose to stick with this famous example.

19. Here we assume that all names in the above sentences are what Kripke (1980) calls 'rigid designators'.
20. We consider (c) below: it is easier to make our point without the difficulties surrounding mental terms.
21. Or otherwise, given any appropriate understanding of the notion of cause, including a deflationary notion. See Ben-Menahem (2018) on causation in physics.

Disclosure Statement

No potential conflict of interest was reported by the author(s).

Funding

This work was supported by Israel Science Foundation [grant number 1148/18], [grant number 690/21].

References

Badash, Lawrence. 1972. "The Completeness of Nineteenth-Century Science." *Isis* 63 (1): 48–58.

Bechler, Zev. 1993. *Three Copernican Revolutions*. Haifa: University of Haifa Press.

Ben-Menahem, Yemima. 2018. *Causation in Science*. Princeton, NJ: Princeton University Press.

Bogen, James. 2017. "Theory and Observation in Science." In *The Stanford Encyclopedia of Philosophy (Summer 2020 Edition)*, edited by Edward N. Zalta. https://plato.stanford.edu/archives/sum2020/entries/science-theory-observation/.

Bohm, D. 1952. "A Suggested Interpretation of the Quantum Theory in Terms of 'Hidden' Variables, I and II." *Physical Review* 85: 166–193.

Bokulich, Peter. 2011. "Hempel's Dilemma and Domains of Physics." *Analysis* 71 (4): 646–651.

Callender, Craig. 1999. "Reducing Thermodynamics to Statistical Mechanics: The Case of Entropy." *The Journal of Philosophy* XCVI: 348–373.

Chalmers, David. 1996. *The Conscious Mind: In Search of a Fundamental Theory*. New York: Oxford Universtiy Press.

Chalmers, J. David. 2013. "How Can we Construct a Science of Consciousness?" *Annals of the New York Academy of Sciences* 1303/1: 25–35.

Chalmers, J. David, and Kelvin McQueen. Forthcoming. "Consciousness and the Collapse of the Wave Function." In *Consciousness and Quantum Mechanics*, edited by S. Gao. Oxford University Press.

Chomsky, Noam. 1968. *Language and Mind*. New York: Harcourt Brace and World.

Chomsky, Noam. 2000. *New Horizons in the Study of Language and Mind*. Cambridge: Cambridge University Press.

Chomsky, Noam. 2003. "Replies." In *Chomsky and his Critics*, edited by L. Antony and N. Hornstein, 255–328. Oxford: Blackwell.

Churchland, Paul. 1996. *The Engine of Reason, the Seat of the Soul*. Cambridge, MA: Bradford and MIT Press.

Crane, Tim, and Hugh D Mellor. 1990. "There is no Question of Physicalism." *Mind; A Quarterly Review of Psychology and Philosophy* 99: 185–206.

Crook, Seth, and Carl Gillett. 2001. "Why Physics Alone Cannot Define the 'Physical': Materialism, Metaphysics, and the Formulation of Physicalism." *Canadian Journal of Philosophy* 31 (3): 333–359. Accessed October 11, 2020. http://www.jstor.org/stable/40232121.

Davidson, Donald. 1970. "Mental Events." In *Essays on Actions and Events*, edited by Donald Davidson, 207–227. Berkeley, CA: University of California Press.

Dennett, Daniel. 2018. *Magic, Illusions, and Zombies: An Exchange*. New York: The New York Review of Books.
Dennett, Daniel, and Marcel Kinsbourne. 1992. "Time and the Observer: The Where and When of Consciousness in the Brain." *Behavioral and Brain Sciences* 15 (2): 183–247.
Dowell, Janice L. 2006a. "Formulating Physicalism." *Philosophical Studies* 131/1.
Dowell, Janice L. 2006b. "Formulating the Thesis of Physicalism: An Introduction." *Philosophical Studies* 131: 1–23. doi:10.1007/s11098-006-6641-y.
Eddington, Arthur S. 1935. *The Nature of the Physical World. London: Everyman's Library*. London: J. M. Dent.
Einstein, Albert. 1905a. "On the Electrodynamics of Moving Bodies." *Annalen der Physik* 17 (10): 895–921. English translation in Einstein, *The Principle of Relativity*. London: Methuen, 1923.
Einstein, Albert. 1905b. "On the Movement of Small Particles Suspended in a Stationary Liquid Demanded by the Molecular-Kinetic Theory of Heat." *Annalen der Physik* 17: 549–560. English translation in Einstein, Investigation on the Theory of the Brownian Movement. London: Dover, 1956.
Einstein, Albert. 1905c. "On a Heuristic Viewpoint Concerning the Production and Transformation of Light." *Annalen der Physik* 17: 132–148. English translation: https://einsteinpapers.press.princeton.edu/vol2-trans/100.
Einstein, A. 1919. "Time, Space, and Gravitation," *Times London*, November 28, 1919, 13–14. Reprinted as "What is the Theory of Relativity?" In Einstein, *Ideas and Opinions*, 227–232. New York: Bonanza Books, 1954.
Einstein, Albert. 1970. "Autobiographical Notes." In *Albert Einstein: Philosopher-Scientist, vol. 2*, edited by P. A. Schilpp. Cambridge: Cambridge University Press.
Feigl, Herbert. 1969. "The Origin and Spirit of Logical Positivism." In *The Legacy of Logical Positivism*, edited by Peter Achinstein, and Stephen F. Barker, 21–38. Baltimore: Johns Hopkins University Press.
Fermi, E. 1936. *Thermodynamics*. New York: Dove, 1956.
Fine, Arthur I. 1984. "The Natural Ontological Attitude." In *Scientific Realism*, edited by J. Leplin, 261–277. Oakland: University of California Press.
Fodor, Jerry. 1974. "Special Sciences (or: The Disunity of Science as a Working Hypothesis)." *Synthese* 28: 97–115.
Fodor, Jerry. 1997. "Special Sciences: Still Autonomous After All These Years." *Nous (detroit, Mich)* 31: 149–163.
Ghirardi, G. C., A. Rimini, and T. Weber. 1986. "Unified Dynamics for Microscopic and Macroscopic Systems." *Physical Review D* 34 (2): 470–491.
Gillett, Carl, and Gene D. Witmer. 2001. "A Physical Need: Physicalism and the *via Negativa*." *Analysis* 61: 302–308.
Giovanelli, Marco. 2020. "'Like Thermodynamics Before Boltzmann.' On the Emergence of Einstein's Distinction Between Constructive and Principle Theories." *Studies in History and Philosophy of Moderns Physics* 71: 118–157.
Graziano, Michael. 2019. *Rethinking Consciousness: A Scientific Theory of Subjective Experience*. New York: Norton.
Hameroff, S., and R. Penrose. 2014. "Consciousness in the Universe: A Review of the 'Orch OR' Theory." *Physicas Life Review* 11 (1): 39–78.
Harlow, Daniel. 2018. "TASI Lectures on the Emergence of Bulk Physics in AdS/CFT." arXiv:1802.01040 [hep-th].
Hawking, Stephen. 1980. "Is the end in sight for theoretical physics?" Inaugural lecture on taking up the Lucasian Professorship of Mathematics in Cambridge University. Printed in: Black Holes and Baby Universes, 42–62.
Hawking, Stephen. 2002. "Godel and the End of Physics." http://yclept.ucdavis.edu/course/215c.S17/TEX/GodelAndEndOfPhysics.pdf.
Hellman, Geoffrey. 1985. "Determination and Logical Truth." *The Journal of Philosophy* 82 (11): 607–616. doi:10.2307/2026415.

Hemmo, Meir, and Shenker Orly. 2016. “Maxwell’s Demon.” In *Oxford Handbooks Online in Philosophy*. Oxford University Press. doi:10.1093/oxfordhb/9780199935314.013.63.

Hemmo, Meir, and Orly Shenker. 2012. *The Road to Maxwell’s Demon: Conceptual Foundations of Statistical Mechanics*. Cambridge: Cambridge University Press.

Hemmo, Meir, and Orly Shenker. 2019. “Two Kinds of High-Level Probability.” *The Monist* 102 (4): 458–477.

Hemmo, Meir, and Orly Shenker. 2020. “Why Functionalism is ‘Token-Dualism’.” PhiSci Archive: http://philsci-archive.pitt.edu/18073/.

Hemmo, Meir, and Orly Shenker. 2021. “A Dilemma for Davidson’s Anomalous Monism”.

Hemmo, Meir, and Orly Shenker. Forthcoming. “Is the Mentaculus the Best System of Our World?” In *Rethinking the Concept of Laws of Nature*, edited by Yemima Ben Menahem. Dordrecht: Springer.

Hempel, Carl. 1969. “Reduction: Ontological and Linguistic Facets.” In *Philosophy, Science, and Method: Essays in Honor of Ernest Nagel*, edited by Patrick Suppes, Sidney Morgenbesser, and Morgan White, 179–199. New York: St. Martin’s.

Hempel, Carl. 1980. “Comments on Goodman’s *Ways of Worldmaking*.” *Synthese* 45: 139–199.

Hoefer, Carl. 2018. “Chap. 2: Scientific Realism Without the Quantum.” In *Scientific Realism and the Quantum*. Oxford: Oxford University Press.

Hoefer, Carl, and Genoveva Marti. 2020. “Water has a Microstructural Essence After all.” *European Journal for Philosophy of Science* 10: 38.

Howard, Don A., and Marco Giovanelli. 2019. “Einstein’s Philosophy of Science”. Stanford Encyclopedia of Philosophy. https://plato.stanford.edu/entries/einstein-philscience/.

Kelvin. 1901. “Nineteenth Century Clouds Over the Dynamical Theory of Heat and Light.” *Philosophical Magazine*.

Kripke, Saul. 1980. *Naming and Necessity*. Cambridge, MA: Harvard University Press.

Lewis, David. 1986. *Philosophical Papers, Volume II*. Oxford: Oxford University Press.

Lewis, David. 1994. “Reduction of Mind.” In *A Companion to the Philosophy of Mind*, edited by Samuel Guttenplan, 412–431. Oxford: Blackwell.

Loewer, Barry. Forthcoming. “The Package Deal Account of Laws and Properties (PDA).” *Synthese*.

Maimon, Amber, and Meir Hemmo. Forthcoming. “Does Neuroplasticity Support the Hypothesis of Multiple Realizability.” *Philosophy of Science*.

Melnyk, Andrew. 1997. “How To Keep the ‘Physical’ in Physicalism.” *The Journal of Philosophy* 94: 622–637.

Melnyk, Andrew. 2003. *A Physicalist Manifesto: Thoroughly Modern Materialism*. Cambridge: Cambridge University Press.

Michelson, A. A. 1903. *Light Waves and Their Uses*. Chicago: University of Chicago Press (quoted in: P.J.E. Peebles “Nineteenth and Twentieth Century Clouds Over the Twenty-First Century Virtual Observatory”).

Montero, Barbara G. 1999. “The Body Problem.” *Noûs* 33 (2): 183–200.

Montero, Barbara G., and David Papineau. 2005. “A Defense of the Via Negativa Argument for Physicalism.” *Analysis* 65 (3): 233–237.

Morganti, Matteo. 2020. “Fundamentality in metaphysics and the philosophy of physics. Part I: Metaphysics.” *Philosophy Compass* 15 (7); Part II: The Philosophy of Physics, *Philosophy Compass* 15. doi:10.1111/phc3.12703.

Ney, Alyssa. 2008a. “Defining Physicalism.” *Philosophy Compass* 3/5: 1033–1048.

Ney, Alyssa. 2008b. “Physicalism as an Attitude.” *Philosophical Studies* 138: 1–15.

O’Connor, Timothy. 2020. “Emergent Properties.” In *The Stanford Encyclopedia of Philosophy (Fall 2020 Edition)*, edited by Edward N. Zalta. https://plato.stanford.edu/archives/fall2020/entries/properties-emergent/.

Peebles, Phillip James Edwin. 2005. “Open Problems in Cosmology.” *Nuclear Physics B (Proc. Suppl* 138: 5–9.

Pineda, David. 2006. “A Mereological Characterization of Physicalism.” *International Studies in the Philosophy of Science* 20: 243–266.

Poland, Jeffrey. 1994. *Physicalism*. Oxford: Clarendon.
Poland, Jeffrey. 2003. “Chomsky’s Challenge to Physicalism.” In *Chomsky and His Critics*, edited by L. Antony, and N. Hornstein, 29–48. Oxford: Blackwell.
Polger, Thomas, W. Shapiro, and A. Lawrence. 2016. *The Multiple Realization Book*. Oxford: Oxford University Press.
Prelević, D. 2018. “Physicalism as a Research Programme.” *Grazer Philosophische Studien* 95 (1): 15–33. doi:10.1163/18756735-000023.
Putnam, Hilary. 1967. “Psychological Predicates.” In *Art, Mind and Religion*, edited by William H. Capitan, and Daniel D. Merrill, 37–48. Pittsburgh, PA: University of Pittsburgh Press. Reprinted as “The Nature of Mental States.” In *Mind, Language and Reality*, edited by H. Putnam, 429–440, Cambridge: Cambridge University Press (1975).
Robinson, Howard. 2020. “Dualism.” In *The Stanford Encyclopedia of Philosophy (Fall 2020 Edition)*, edited by Edward N. Zalta. https://plato.stanford.edu/archives/fall2020/entries/dualism/.
Searle, John. 1997. *The Mystery of Consciousness*. New York: The New York Review Books.
Shapiro, Lawrence. 2018. “Reduction Redux.” *Studies in History and Philosophy of Science* 68: 10–19.
Shenker, Orly. 2015. “Davidsonian Descriptions as a Principle Theory.” *Iyyun: The Jerusalem Philosophical Quarterly* 64: 171–190. [in Hebrew].
Shenker, Orly. 2017. “Flat Physicalism: Some Implications.” *Iyyun, the Jerusalem Philosophical Quarterly* 6: 1–15.
Shenker, Orly. 2020. “Denialism: What do the so-Called ‘Consciousness Deniers’ Deny?” *Iyyun: The Jerusalem Philosophical Quarterly* 68: 307–337.
Smart, Jack J.C. 1978. “The Content of Physicalism.” *The Philosophical Quarterly* 28: 339–341.
Smolin, L. 2006. *The Trouble with Physics*. Boston, MA: Houghton Millfin.
Spurrett, D., and David Papineau. 1999. “A Note on the Completeness of ‘Physics’.” *Analysis* 59/1: 25–29.
Stoljar, Daniel. 2010. *Physicalism*. London: Routledge.
Stoljar, Daniel. 2017. “Physicalism.” In *The Stanford Encyclopedia of Philosophy (Winter 2017 Edition)*, edited by Edward N. Zalta. https://plato.stanford.edu/archives/win2017/entries/physicalism/.
Uffink, Jos. 2001. “Bluff Your Way in the Second Law of Thermodynamics.” *Studies in History and Philosophy of Science Part B - Studies in History and Philosophy of Modern Physics* 32 (3): 305–394.
van Fraassen, Bastain C. 2002. *The Empirical Stance*. New Haven, CT: Yale University.
van Riel, Raphael, and Robert Van Gulick. 2019. “Scientific Reduction.” In *The Stanford Encyclopedia of Philosophy (Spring 2019 Edition)*, edited by Edward N. Zalta. https://plato.stanford.edu/archives/spr2019/entries/scientific-reduction/.
van Roojen, Mark. 2018. *“Moral Cognitivism vs. Non-Cognitivism.” In The Stanford Encyclopedia of Philosophy (Fall 2018 Edition)*, edited by Edward N. Zalta. https://plato.stanford.edu/archives/fall2018/entries/moral-cognitivism/.
Vintiadis, Elly, and Constantinos Mekios. 2018. *Brute Facts*. Oxford, UK: Oxford Scholarship Online.
von Neumann, J. 1932. *Mathematical Foundations of Quantum Mechanics (trans. R. T. Beyer 1955)*. Princeton: Princeton University Press.
Wallace, D. M. 2012. *The Emergent Multiverse*. Oxford: Oxford University Press.
Wilson, Jessica. 2006. “On Characterizing the Physical.” *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition* 131 (1): 61–99.
Worley, Sara. 2006. “Physicalism and the Via Negativa.” *Philosophical Studies* 131 (1): 101–126.